

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2003, 19:24:17 ; Search time 2554 Seconds
(without alignments)
10608.868 Million cell updates/sec

Title: US-09-997-900-1
Perfect score: 1673
Sequence: 1 gctcttcagcagcaaaaa.....tgatttggcctcatttcg 1673

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	565.4	33.8	1380	11	AY105043
2	502.4	30.0	9	9	AI992657
3	494.2	29.5	605	10	AV439911
4	404.6	24.2	508	14	N96256
5	399	23.8	423	10	AV801585
6	392	23.4	802	14	BQ515774

Result No.	Score	Query Match	Length	ID	Description
7	376.8	22.5	454	14	T88501
8	372.4	22.3	418	10	AV796666
9	366.4	21.9	679	14	BO407261
10	366	21.9	714	14	BO863607
11	359.2	21.5	431	10	AV797506
12	346.2	20.7	422	10	AV816962
13	338.2	20.2	410	10	AV805399
14	334.6	20.0	607	10	AM10049
15	333.2	19.9	675	10	AM098716
16	332	19.8	631	10	AM585871
17	328.2	19.6	508	9	AA394708
18	325.6	19.5	782	12	BG581436
19	323.8	19.4	541	12	BE999544
20	322	19.2	618	14	BO913769
21	322	19.2	797	12	BG836929
22	321.6	19.2	538	10	AM508667
23	311	18.6	504	12	BG507593
24	310.6	18.6	638	12	BE823545
25	309.8	18.5	632	14	BO415906
26	299.6	17.9	409	10	AV808789
27	296.4	17.7	601	10	AM596137
28	296	17.7	624	10	AM181183
29	294.4	17.6	688	13	BJ257372
30	288.8	17.3	552	13	BJ310586
31	283	16.9	382	10	AV790195
32	282	16.9	674	13	BI931584
33	279	16.7	555	10	BE433354
34	277.2	16.6	618	10	AV916746
35	275.2	16.4	653	12	BF481461
36	272.6	16.3	388	10	AV804485
37	272.2	16.3	631	13	BJ462897
38	272	16.3	466	10	BE346767
39	266.2	15.9	395	10	AV817239
40	266	15.9	417	17	BH847386
41	262.6	15.7	517	12	BF068113
42	258.8	15.5	660	10	AV922042
43	255.8	15.3	505	14	BO629151
44	255.6	15.3	722	12	BG126660
45	252.6	15.1	795	14	BQ515775

ALIGNMENTS

RESULT 1
AY105043
LOCUS
DEFINITION
Zea mays
ACCESSION
AY105043
VERSION
AY105043.1 GI:21208121
KEYWORDS
HTC.
SOURCE
ORGANISM
Zea mays.
Zea mays.
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
AUTHORS
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE
Maize Mapping Project/Dupont Consensus Sequences for Design of
Overgo Probes
JOURNAL
Unpublished (2002)
REFERENCE
2 (bases 1 to 1380)
AUTHORS
Coe,E.C.
TITLE
Direct Submission
JOURNAL
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
FEATURES
source
1..1380
/organism="Zea mays"
/db_xref="MaizeDB:633742"
/db_xref="taxon:4577"
/clone="PC0069007"

/clone_1lb="Waize Mapping Project/DuPont Cornsensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed Dupont contigs; this resource was
assembled by Dupont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

BASE COUNT 365 a 294 c 342 g 377 t 2 others
ORIGIN

Query Match 33.8%; Score 565.4; DB 11; Length 1380;
Best Local Similarity 70.2%; Pred. No. 1.2e-164;
Matches 788; Conservative 0; Mismatches 331; Indels 4; Gaps 2;

385 AGAGTCTTGCTGTTGCTGTCAGAGAGCAAGGCTTATTCACCATGTTCTGTGAA 444
11 AGTCGCTCGCCGCTGAGGCTCAACAGAGCAAGGCCCTTCCACCATTTGCTCTCCGGA 70

445 CTGAAGGCTACTTTCAGCAGGCTATCGCAACTCTCAAGGCTCTTATGTTCTAAAG 504
71 CCGAGAGGCTGCTCAACCAAGTATCGAGCAGCTCATAGCTTGTCAAGTCTCTGAGTG 130

505 TTGAAGTATCTCAAGTACAGCCCAAGTGAAGCTGAGCTGATGCTTTGTAAGTGAATG 564
131 TGGAGATCTATCTAAAGAACCTCAGGTTGAAGAGAGAGCTGATGCTTATAAACTAATG 190

565 CACATCCAGATCCAGGAGATCATGTGGCTGAGTGCACATTCGAGCAAGCAAGTTTG 624
191 TTGAACCTGATCAGCCCTGAGGCTCATGTTTATGTTGATTTTTCAGCAAAATTTG 250

625 TAGATATAGCGGAACATGATTCGATTCGAGGATGAGATCCTGCAAAATGATTTG 684
251 TTGATATATCGAAGAAACCTTACATGAGAGTGTGAGATCCCGCAAAATTTCTG 310

685 CTGTAGAAAGAAATTTGAAAAGTTTCAGATCAGAGAGATGTTAGAGCAAGAAATAG 744
311 CAGTGCAGAGAAATCTATGGAATTTGGAATCAAAAGAAATTTGAGAGCAAGAAATTTG 370

745 CACTGAGAGGGAAGATGAGTGGGCTGCTCATTTTGGCGATTTTCAGCAAGCATCT 804
371 CTTTGAGACGTGAAAGATGTTGTCAGACGCCCTTCTGCGGATTTTGTGCTCTCTT 430

805 ATCCAGATCTCAAGAGCAAGCGCTGTTAGTGTTCGAGATGCAAAAAGAGCA 864
431 ACCAGACGTGATAGAGCGCTTGCCGAAAATCCACTTACATCTGTAATAGAGCACTTA 490

865 TTGTCCCTCAAAAAGAAACATC--AGCAGGGGAGATGTTTATCCCGTTGAGCCATTTT 921
491 ATGGCAGTTTGTGCAACCATCCAAATGCTGGGGGTGATGTTTATCCGTGGAATCTTACG 550

922 TTGACCCCAAGATACATGATTCGAGGCTCAGTGGGATCTTCACGAGCAAGATA 981
551 AGAGCTTATAGTGAACCATGATGATGATGCTCATTTGGGGTGTCTGGATGATGATGATG 610

982 CGAGTGAAGTACGCTGCTACATC--TCTATCATTGCTTGTGTAATGATATTCAGAGATCTT 1040
611 CGAGTGAAGTACGCTGCTACATCCTCTCATCTGTGATGATGATGATGATGATGATG 670

1041 AATATTTGCTGCTGTTTCTGCTCGAAGGGATCAATATCCAGAGCTTGGCCGTAGGA 1100
671 AATATTTGCTGCTGTTTCTGCTCGAAGGGATCAATATCCAGAGCTTGGCCGTAGGA 730

1101 CATGCTGAACCAAGGCAATTTTCAGCATATACAAAGATTTACTGCAACAGATGATTCG 1160
731 CCAAGCTGAGAGGAAAGCAATTTTCCGGATTAACAACATTTTCTGCTGATGTTGATTC 790

1161 GTGAGCAAAATTTGTCAGCAACTTTACAACCTGATAGATGATGATGATGATGATGAT 1220
791 ATTGAGAGATTTAGTTGAGCAAGCTTTTACAACCTTATGATGATGATGATGATGATGAT 850

1221 ACTGATTTGCTGCTGTTTCTGGAAGAGAACTGATGATGATGATGATGATGATGATGAT 1280

Db 851 ACCCACTACCTTTTGTCTGAAGGAACTTATGCTTATTAAGTTTCTGTCACACTGCT 910

1281 GCTAGAGAGATGTCCTGAGCATTCATGATATTTTCAGGGCTTAAGCTGTGACGATCT 1340

911 GCTCGAGAGAAATCTTACATGATTTGCTGAATCTTCCGAGCAAAACCTGTTGATTTCT 970

1341 GATCACAATTAATTTGCACTTACTGAGGATCTAGACAAGATGTTGACATGCAAGAG 1400

971 GACCACACAGTAAGCCTTACAGCTTACTGAGATCTTACAAAGATGTTGACATCAAGAG 1030

1401 TTATGAGAGCCCTATGATATGATGAGCTTGCAGAAACCGCTGTCGATTTGCTCT 1460

1031 TTATGAGAGCCCTATGATATGATGAGCTTGCAGAAACCGCTGTCGATTTGCTCT 1090

1461 GAATCGGAGTGCAGCTCAAGTACCTGTCGATATCTCTTTC 1503

1091 GAATCGAAGGTGACATCCAGTACCTCGCGCTACTCTCTTC 1133

RESULT 2
A1992657 539 bp mRNA linear EST 08-SEP-1999
LOCUS A1992657
DEFINITION 701558592 A. thaliana, Ohio State clone set Arabidopsis thaliana
ACCESSION C1992657
VERSION A1992657.1 GI:5839562
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 539)
REFERENCE
AUTHORS Chen, J., Momiyama, M., Chan, E., Mooney, M., Carreon, B., Gilliland, D.,
Wang, X., Hillman, J., Guebler, K., Kim, C., Doyle, M., Brzozka, P.,
Gorgone, G., Burns, D., Griffin, D., Mounoudou, M., Nguyen, D., Tan, R.,
Rose, M., Warren, B., Ton, B., Kaslury, K., Borillo, C., Carpio, P.,
Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobrigha, A., Murry, L.,
Turner, C., Krikorian, S., Elder, L., and Hanson, D.
TITLE Arabidopsis thaliana Gene Expression Microarray
JOURNAL Unpublished (1999)
COMMENT Contact: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte
Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3824
Email: service@genomesystems.com.
FEATURES
source
1. 539
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="701558592"
/note="cdna library was made from selected clones from the
Arabidopsis thaliana Ohio State clone set."
BASE COUNT 157 a 124 c 105 g 153 t

Query Match 30.0%; Score 502.4; DB 9; Length 539;
Best Local Similarity 98.1%; Pred. No. 3.1e-145;
Matches 530; Conservative 0; Mismatches 6; Indels 4; Gaps 2;

1125 CGCATTCACACATGTTATCTGCAACAGATGATGCTGTCAGCAAAATTTGTCGACAACTT 1184

539 CGCATTCACACATGTTATCTGCAACAGATGATGCTGTCAGCAAAATTTGTCGACAACTT 481

1185 TACAACCTGATAGTGCATGAGGTCAGATGATGATGATGATGATGATGATGATGATG 1244

480 TACAACCTGATAGTGCATGAGGTCAGATGATGATGATGATGATGATGATGATGATG 421

1245 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1304

```

Db      420 GAAGTGAATGCTGATTAAGATTGCCGTGAACGCTGCTGCTAGAGAAGATGTCCTGGACATT 361
OY      1305 GCATGATATTTTCAGGGCTAAAGCTGTGTGACGTATGATGCACACAACTTACTTTGACGCTT 1364
Db      360 GCATGATATTTTCAGGGCTAAAGCTGTGTGACGTATGATGCACACAACTTACTTTGACGCTT 301
OY      1365 ACTGGGAGATCTAGACAAGATGCTTGGACCTGCAAGGTTATTGGAGCCCTATGTTATATGT 1424
Db      300 ACTGGGAGATCTAGACAAGATGCTTGGACCTGCAAGGTTATTGGAGCCCTATGTTATATGT 241
OY      1425 GAGGTTCGAAGAACCAGCTGCTGTGCGATTGGCTTCGATGCGAGTGCAGCTCCAAAGTAC 1484
Db      240 GAGGTTCGAAGAACCAGCTGCTGTGCGATTGGCTTCGATGCGAGTGCAGCTCCAAAGTAC 181
OY      1485 CTTCGTGATGATCTCTCTCTTTTAAACAGGCTAACCGTTGCAGAGTGCATCCATCGACACA 1544
Db      180 CTTCGTGATGATCTCTCTCTTTTAAACAGGCTAACCGTTGCAGAGTGCATCCATCGACACA 121
OY      1545 TCAGAAACTTTGGAAGCTTAAAGCTTTCATTACACAGCTCATGTAAC--CTCAAGAAGACA 1601
Db      120 TCAGAAACTTTGGAAGCTTAAAGCTTTCATTACACAGCTCATGTAACCAACCAAGAAGACA 61
OY      1602 CAGAGAGACTGCGTCATATATGTTGTGACTTTGTTTATGAACAATTAGCTGATTTTG 1661
Db      60 CAGAGAGACTGCGTCATATATGTTGTGACTTTGTTTATGAACAATTAGCTGATTTTG 1

RESULT 3
AV439911/c 605 bp mRNA linear EST 14-NOV-2000
LOCUS      AV439911 Arabidopsis thaliana above-ground organ two to six-week
DEFINITION Old Arabidopsis thaliana cDNA clone APD28a06_f_3', mRNA sequence.
ACCESSION  AV439911
VERSION     AV439911.1 GI:7610260
KEYWORDS   EST.
SOURCE     thale cress.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 605)
            Asanizu, E., Nakamura, Y., Sato, S. and Tabata, S.
            A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
            DNA Res. 7, 175-180 (2000)
JOURNAL    20363093
MEDLINE    Contact: Erika Asanizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asanizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES   Location/Qualifiers
            source          1..605
                        /organism="Arabidopsis thaliana"
                        /strain="Columbia"
                        /db_xref="taxon:3702"
                        /clone="APD28a06_f"
                        /clone_lib="Arabidopsis thaliana above-ground organ two to
                        six-week old"
                        /tissue_type="aboveground organs"
                        /dev_stage="two to six-week old"
                        /note="Vector: pBluescriptII SK-, site_1: EcoRI; site_2:
                        XhoI"
BASE COUNT 193 a 135 c 110 g 167 t
ORIGIN
Query Match 29.5%; Score 494.2; DB 10; Length 605;
Best Local Similarity 98.8%; Pred. No. 1,2e-142;
Matches 509; Conservative 0; Mismatches 3; Indels 3; Gaps 1;
OY 1151 AGATGAATCGCTGACCAATTTGCTGACGACAACTTAGATGTCATGAGGT 1210
|||||

```

```

Db      605 AGATGAATCGCTGACCAATTTGCTGACGACAACTTAGATGTCATGAGGT 546
OY      1211 CCATGATCTTACTCATTTGGCATTTTCTGAAAGAGATGCTGATTAAGTTGCCGT 1270
Db      545 CCATGATCTTACTCATTTGGCATTTTCTGAAAGAGATGCTGATTAAGTTGCCGT 486
OY      1271 GAACGCTGCTGCTAGAGAAGATGTCCTGACATTGGCTAGATTTTCAGGCTAAAGCTGT 1330
Db      485 GAACGCTGCTGCTAGAGAAGATGTCCTGACATTGGCTAGATTTTCAGGCTAAAGCTGT 426
OY      1331 TGACGTATCTGATACACAACTTATCTTGGACGTTACTGGGATCTAGACAAGATGTTGC 1390
Db      425 TGACGTATCTGATACACAACTTATCTTGGACGTTACTGGGATCTAGACAAGATGTTGC 366
OY      1391 ACTGCAAGGTTATTGGAGCCCTATGTTATATGATGAGCTTGCAGAAACCGCTGTCGCG 1450
Db      365 ACTGCAAGGTTATTGGAGCCCTATGTTATATGATGAGCTTGCAGAAACCGCTGTCGCG 306
OY      1451 ATTGCTGCTGGAATTCGGAGTGGACTTCCAAAGTACCTTCGTGATGATCTCTTCTTTTAA 1510
Db      305 ATTGCTGCTGGAATTCGGAGTGGACTTCCAAAGTACCTTCGTGATGATCTCTTCTTTAA 246
OY      1511 AGGCTAAACCGTTGCAGAGTGCATTCATGCAATCAGAAACTTTGGAGGTAAGTTT 1570
Db      245 AGGCTAAACCGTTGCAGAGTGCATTCATGCAATCAGAAACTTTGGAGGTAAGTTT 186
OY      1571 CATTACACAGCTCATGTAAC--CTCAAGACAGACAGAGACTGCGTCATATATGTTT 1627
Db      185 CATTACACAGCTCATGTAACCAACCTCAAGACAGAGAGACTGCGTCATATATGTTT 126
OY      1628 GTGACTTTGTTTATGAACAATTAGCTGATTTTG 1662
Db      125 GTGACTTTGTTTATGAACAATTAGCTGATTTTG 91

RESULT 4
N96256 508 bp mRNA linear EST 05-JAN-1998
LOCUS      N96256 Lambda-PRL1 Arabidopsis thaliana cDNA clone G7C11T7, mRNA
DEFINITION 21855 Lambda-PRL1 Arabidopsis thaliana cDNA clone G7C11T7, mRNA
ACCESSION  N96256
VERSION     N96256.1 GI:2748591
KEYWORDS   EST.
SOURCE     thale cress.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 508)
            Newman, T., deBruin, F.J., Green, P., Keegstra, K., Kende, H., McIntosh
            L., Ohlroge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel
            E. and Somerville, C.
            Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
            Plant Physiol. 106, 1241-1255 (1994)
JOURNAL    95148729
MEDLINE    Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
            Lansing, MI
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 223j3tcn@dm.cl.msu.edu
            Seq primer: T7.
FEATURES   Location/Qualifiers
            source          1..508
                        /organism="Arabidopsis thaliana"
                        /strain="var Columbia"
                        /db_xref="taxon:3702"
                        /clone="G7C11T7"
                        /clone_lib="Lambda-PRL1"

```


ACCESSION B0515774
 VERSION B0515774.1 GI:21374643
 KEYWORDS EST.
 SOURCE potato.
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 REFERENCE Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C.,
 Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and
 Karameycheva, S.A.
 Generation of a set of potato cDNA clones for microarray analyses
 Unpublished (2002)
 Other ESTs: EST623190
 Contact: Robin Buell
 TITLE The Institute for Genomic Research
 JOURNAL 9712 Medical Center Dr, Rockville, MD 20850, USA
 COMMENT Email: potato@ligr.org
 This clone is available through the Research Genetics, contact the
 Research Genetics for further information 1-800-711-6195 or
 cda@resgen.com
 Seq primer: 13.
 FEATURES
 source Location/Qualifiers
 1..802
 /organism="Solanum tuberosum"
 /cultivar="Kennebec or Binje"
 /db_xref="taxon:4113"
 /clone="STMU02"
 /clone_lib="Generation of a set of potato cDNA clones for
 microarray analyses mixed potato tissues"
 /tissue_type="mixed tissues"
 /lab_host="SOLR"
 /note="vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
 XhoI; supplier: Combination of untreated and Phytophthora
 infestans-treated libraries of stolons, leaves, leaflets,
 axillary buds of stem explants, petioles, germinating eyes
 , tubers, or roots."
 BASE COUNT 234 a 145 c 201 g 222 t
 ORIGIN
 Query Match 23.4%; Score 392; DB 14; Length 802;
 Best Local Similarity 68.9%; Pred. No. 1.3e-110;
 Matches 555; Conservative 0; Mismatches 245; Indels 6; Gaps 1;
 QY 544 TCATGCTTTAAAGATGATGCATCCAGATCCAGGCGAGATGCTGCTGTTG 603
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 2 TGAATGCTCAATAAAGTGAATGCAGATGCAACCAAAAGCTGAATCATGCTGTTG 61
 QY 604 ACACATTCAGAGCAAGATGTGATATAGCGGACATGCATGACTATCGAGTAACTG 663
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 62 ACATCTTCAGAGAAATTTGGATATATTCAGACACCTTTCTGACTATTTGAGTAACTG 121
 QY 664 GAGATCCTGAAAAATGATTTGCTGTAAGAAAGTAATTTGAAAAAGTTTCAGATCAGAGAGA 723
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 122 GTGATCCAGGAAATAGGCTGCGCTCTGAGAAATTTAAGCAAGTTGGATAGGAGAAC 181
 QY 724 TTGTAAAGCAGAGAAAGATAGCACTGAGAAAGGAAAAAGATGCTGCACTCTCATTTT 783
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 182 TAGCCGGAACGTGGAGATGCTTTAGAGACGAGAAAAAGAGTGGTGAAGATGCTCTTTT 241
 QY 784 GGGGATTTTCAGAGCATCTATCCAGATCTCAGAGAGAGGCGCTGTAGTGTCTTC 843
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 242 GGAAGTTTTCGACGCTTTCTATTCAGATCTCGAAGAAAAAATCTCCACGTA-----TTC 295
 QY 844 GAAGTAGCAAAAAAGAGCATTTGCTCCCAAAAGAAACATCAGCAGGAGGAGATGTTT 903
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 296 TGAACAACAAGAGTGCACAAATGACATGAGACCGAGATGATGATGGGATGATGTTT 355
 QY 904 ATGCCCTTGAGCATTTTGGACCCAGAGTACATGTAATTCGACCGCTCATCTGGGAGAC 963
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 356 ATCCCTGTGAATCCAGTGAATCATCTCAGTTAATCAAAATTCGATGCCACTGGGGAT 415

QY 964 TTCTACAGCAGAGATAGAGTGAAGTGCATGCGCTGATCTATCATCTTGAATG 1023
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 416 ATCTGATATGAGATTTGTCTGGATTCGATTCACACATTAAGCATGCTTGCAACA 475
 QY 1024 ATATTCCAGAGTTCTTAATATTGTCAGTGTGTTTTCCTCGAAGGGATACATATCC 1083
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 476 ATGCTCTCTGGGGCCCTAAATTTGATGAAGTGCTGTATATCTGACAGAGGCTAATAGTTC 535
 QY 1084 AGAGCTTGCCCGTAGGACATCTGGAACCAAGGGCATTTTCAGCATTAACAGATTATAC 1143
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 536 AGAGTTTCTCTGAGCCCTGACAGACAGAGGGCTTTTCAGATTAATACACAGTGTCC 595
 QY 1144 CTGCACAGATGAAATGCTGACGAATTTGGTGCAGCACTTACAACTCTAGATGTCC 1203
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 596 CTGGAACATGATAAATATAGGAAATTTGTCAGCAATTTAACAAGTTGATAGATGTTCC 655
 QY 1204 ATGAGTCCATGATCTTACTCATTTTCCATTTTTCGAAGAGACTAGTGCATTAAGA 1263
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 656 ATGAGCTTCAGATATATACATTTGCTGAGCGGGAGTAAATGCTATTAAG 715
 QY 1264 TTGCGGTGAAGCTGCTGCTAGAGAGATGCTGACATTTGCTATTTTCAGGGCTA 1323
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 716 TAGCCGCAATGACGACGAGAGGAGATGCTGTGATATTCACATATTTTCCGTCTA 775
 QY 1324 AAGCTGTGACGATCTGATCACACA 1349
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 776 AACCTGTGATGCTGCTGATCATACA 801

RESULT 7
 T88501 454 bp mRNA linear EST 09-JAN-1998
 LOCUS 12197 Lambda-PRL2 Arabidopsis thaliana cDNA clone 157J2377, mRNA
 DEFINITION
 T88501
 T88501
 EST.
 ACCESSION T88501
 VERSION T88501.1 GI:936335
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 454)
 Newman, T., debruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh
 , L., Ohlroge, J., Raikhel, N., Somerville, S., Thomasow, M., Retzel
 , E. and Somerville, C.
 Genes galore: a summary of methods for accessing results from
 large-scale partial sequencing of anonymous Arabidopsis cDNA clones
 Plant Physiol. 106, 1241-1255 (1994)
 95148729
 On Apr 14, 1993 this sequence version replaced gi:716853.
 Contact: Thomas Newman
 MSU-DOE Plant Research Laboratory
 Michigan State University
 MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
 Lansing, MI
 Tel: 517-353-0854
 Fax: 517-353-9168
 Email: 22313tcn@lhm.cl.msu.edu
 Seq primer: T7 dye primer.
 FEATURES
 source Location/Qualifiers
 1..454
 /organism="Arabidopsis thaliana"
 /strain="var Columbia"
 /db_xref="taxon:3702"
 /clone="157J2377"
 /clone_lib="Lambda-PRL2"
 /note="vector: lambda zip-lox; Site.1: Sal; Site.2: Not;
 Lambda PRL2 is a cDNA library derived from equal
 quantities of 4 pools of mRNA. The mRNA sources were 1) 7
 day germinated etiolated seedlings; 2) tissue culture
 grown roots; 3) staged plants half with 24 hour light
 cycle, half on 16 hr light, 8 hour dark- rosettes; 4)

same plant as 3 but aerial tissue (stems, flowers and siliques). The vector is BR1's lambda Zip-Lox. The cDNA inserts were directionally cloned with Sal-I Not arms using oligo dT primed cDNA."

BASE COUNT	124 a	92 c	104 g	118 t	16 others
ORIGIN					

Query Match	22.5%;	Score 376.8;	DB 14;	Length 454;
Best Local Similarity	91.2%;	Pred. No. 5.1e-106;		
Matches 412;	Conservative 0;	Mismatches 38;	Indels 2;	Gaps 2

OY	728	AAGACAGGAAAGATTACACACGAGAAAGGAAAGATGGGTGCAACTGTCATATTTGGCG	787
Db	1	AAGCAGNNMAAGATACACTGAGAAAGGAAAAAGATGGGTGCAACTNCTCCATTTTGGCG	60
OY	788	ATTTTCAGCAGCATCTATCCAGATCTCAAGAGGACAGCGCCCTGTATGATGTTCTTCCAG	847
Db	61	ATTTTNNAGCACATCTTATCCAGATCTCAAGAGGACAGCGCCCTGTATGATGTTCTTCCAG	120
OY	848	TAGCAAAAAAGAGCCATTTCCCTCCAAAAGAAACATACAGAGGGGGAGATGTTATCC	907
Db	121	TAGCAAAAAAGAGCCATTTGTCCTCCAAAAGAAACATACAGAGGGGGAGATGTTATCC	180
OY	908	CGTTGAGCCATTTTATACCCCAAGATACATCGTATTTCTGACGCTCAGTGGGGACTTCT	967
Db	181	CGTTGAGCCATTTTATGACCCCAAGATACATCGTATTTCTGACGCTCAGTGGGGACTTCT	240
OY	968	CACGTGACGAGATACGAGTGGACTACGGTGCATCTATCATCTTGCTGTGAATGAATAT	1027
Db	241	CACGTGACGAGATACGAGTGGACTACGGTGCATCTATCATCTTGCTGTGAATGAATAT	300
OY	1028	TCCAGGAGTTTATATATGTGAGATGGTGTTTTGCGTCGAGAGGGGATACATAATCCAG	1087
Db	301	TCCAGGAGTTTATATATGTGAGATGGTGTTTTGCGTCGAGAGGGGATACATAATCCAG	360
OY	1088	CTTGAGCCGTAGACATCTGTAAGAACCAAGGCGATTTCAGCATTTTACAAGTTATACGTGC	1147
Db	361	CTTGAGCCGTAGACATCTGTAAGAACCAAGGCGATTTCAGGATT - NACAGATTATACGTGC	419
OY	1148	AACAGATGAATCGGTGACGAAATTTGGTGACG	1179
Db	420	AACATTTG - ATCGGTGACGAAATTTGGTGAATG	450

RESULT	8
LOCUS	AV796666/c
DEFINITION	AV796666 418 bp mRNA linear EST 29-MAR-2002 AV796666 RAF1 Arabidopsis thaliana cDNA clone RAFL09-09-C24 3 , mRNA sequence.
ACCESSION	AV796666
VERSION	AV796666
KEYWORDS	AV796666.1 GI:19830649
SOURCE	EST .
ORGANISM	thale cress. Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 418)
REFERENCE	Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh.M., Ishii.Y., Arkawa.T., Shibata,K., Shingawa.A., Muramatsu,M., Hayashizaki,Y. and Shinzekaki,K
AUTHORS	
TITLE	Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL	Unpublished (2002)
COMMENT	Contact : Motoaki Seki

and XhoI was ligated to modified Lambda FIC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified plasmid vector. Please visit our web site (http://www.gsc.riken.go.jp/e/planb/index_e.html) for further details.

FEATURES	Location/Qualifiers
source	1. .418

```

/db_xref="Raxxon:3702"
/clone="RAFL09-09-C24"
/clone_1lb="RAFL9"
/dev_stage="plants at various developmental stages from
germination to mature seeds"
/lab_host="DH10B"
/note="Site_1: BamHI; Site_2: SalI; subjected to
dehydration (1, 2, 5, 10/24 hr) and cold (1, 2, 5, 10, 24
hr) treatments"
126 a 96 c 81 g 115 t

```

Qy 1607 AGACTGCGCATATATGTCTTGTGACCTTTGTTATGAACAATAAGTCATTTTGCC 1664
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 58 AGAGTCGGTAGATATATGTCTTGTGACCTTTGTTATGAACAATAAGTCATTTTGCC 1

RESULT 9	BQ407261	BQ407261	679 bp	mRNA	linear	EST 22-MAY-2002
LOCUS	BQ407261					
DEFINITION	GA_EB0104F04f Gossypium arboreum 7-10 dpa fiber library Gossypium					
	arboreum cDNA clone GA_EB0104F04f, mRNA sequence.					
ACCESSION	BQ407261					
VERSION	BQ407261.1	GI:21094948				
KEYWORDS	EST.					
SOURCE	Gossypium arboreum.					
ORGANISM	Gossypium arboreum					

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 679)
Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
D., Wood, T.C., Leslie, A. and Wilkins T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2000)
Contact: Wing RA

Db 121 ATGTTTTAAAGTTGAGATATCTCAAGGACACAGAGTGAAGTGAATGATGCTCA 180
QY 553 TAAAGTGAATGACATCCAGATTCAGAGGAGAGATCATGTGAGTGAACATTCAC 612
Db 181 TTAAAGTCAATGAGATTCACAGCTACCGATTGAGGTGAGTGGTGGAGATCTTAA 240
QY 613 GAGCAGAGTTGATGATATGACGAGACATGCTATGACTATGAGTGAATCGATCTG 672
Db 241 GAGCTAGGATTTGATGATATTTCAAGAGACTATTAACCATTTGAAGTAACTGGTACCTG 300
QY 673 GAAAGTGAATGCTGTGAGAAAGAAATTTGAAAAAGTTTCAGATCAGAGATTTGAAGA 732
Db 301 GAAAGTGAATGCTGTGAGAAAGAAATTTGAAAAAGTTTGAAGTAACTGGTAAAGGA 360
QY 733 CAGGAAAGATGAGCTGAGAGAGGAGAAAGATGGTGCACTGCTCATTTTGGCGATTTT 792
Db 361 CTGGAAGATGCTGCTTGAAGAGAGAAAAATGGGTGAATGCTCTCTTCTTGAGATTTT 420
QY 793 CAGCAGCATCTATCCAGATCTCAGAGGACAGCGCTGTAGTGTCTTGAAGTAGCA 852
Db 421 CAGCAGCTTCATATCCAGACC-----TTGAAGCTCCCTTCAGACCA 462
QY 853 AAAAGAGGAGCATTTGCCCTCAAAAGAAACATCAGCAGGAGATGTTTATCCGTTG 912
Db 463 AAACAAAACAGAAATATCTGGATCTGATACATCAGTTGGGGAGAGATGTTTATCCAGTTG 522
QY 913 AGCCATTTTGAAGTACCATGATCGTATCTCGACGCTCATCGGGAGCTTCAGATG 972
Db 523 AGCTATATGATGCTATGCTTCAATCAAGTCTTTGATGACATTTGGGGTCTTCAACG 582
QY 973 ACGAAGATACGATGAGTACAGGCTGCATACCTATCATCTTCTTGAATGATATTTCCAG 1032
Db 583 AAGAAGATACACCGGGCTTCGCTCAACACCTTATCAATGATTTGAAGACATCCCTG 642
QY 1033 GAATTTTAAATATGCTGAGTGTGTTTTCGCTGGAAGGGATCAATATTCAGAGCTTG 1092
Db 643 GAGTCTCATATATGAAACAGAGATTTTTCGCTGTAAGGCTCAACATTCAGAGCTTAG 702
QY 1093 CCCTAGACATG 1104
Db 703 CAGTTGCTCATG 714

RESULT 11
AV797506/c 431 bp mRNA linear EST 29-MAR-2002
LOCUS AV797506 RAFL9 Arabidopsis thaliana cDNA clone RAF109-12-G04 3',
DEFINITION mRNA sequence.
ACCESSION AV797506
VERSION AV797506.1 GI:19831489
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 431)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Saitou, M., Nakajima, M.,
Oono, Y., Sekurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.,
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@tc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda FL-C-1 vector (Carninci et

al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified Bluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.

FEATURES
source location/Qualifiers
1..431
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="RAFL09-12-G04"
/clone_11b="RAFL9"
/dev_stage="plants at various developmental stages from
germination to mature seeds"
/lab_host="DH10B"
/note="Site 1: BamHI; Site 2: SalI; subjected to
dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
hr) treatments"

BASE COUNT 136 a 100 c 75 g 120 t

ORIGIN

Query Match 21.5%; Score 359.2; DB 10; Length 431;
Best local similarity 98.4%; Pred. No. 1.6e-100;
Matches 374; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

QY 1286 AAGAGATGCTGAGCATTTGATGATTTTCAAGGCTAAAGCTGTTGAGTATCTGATCA 1345
Db 431 AAGAGATGCTGAGCATTTGATGATTTTCAAGGCTAAAGCTGTTGAGTATCTGATCA 372
QY 1346 CACATTTACTTTGACCTTACTGCGGATCTAGACAAGATGTTGACCTCAAGCTTATT 1405
Db 371 CACATTTACTTTGACCTTACTGCGGATCTAGACAAGATGTTGACCTCAAGCTTATT 312
QY 1406 GGAGCCCTATGATATGAGGTGTCAGAACCGGCTGCTGGCATTTGGCTGTGAATC 1465
Db 311 GGAGCCCTATGATATGAGGTGTCAGAACCGGCTGCTGGCATTTGGCTGTGAATC 252
QY 1466 GGGAGTGAAGCTCAAGTACCTTCGTGATACCTCTTTTAAAGGCTAAAGCTTGC 1525
Db 251 GGGAGTGAAGCTCAAGTACCTTCGTGATACCTCTTTTAAAGGCTAAAGCTTGC 192
QY 1526 AGAGTGCATTCATCGAACAATCGAAGCTTGAAGCTTGAAGCTTGAAGCTTAT 1585
Db 191 AGAGTGCATTCATCGAACAATCGAAGCTTGAAGCTTGAAGCTTGAAGCTTAT 132
QY 1586 GAAC--CTCAAGACAGACAGAGAGAGAGCTGCGATATATGTTGTGACTTTGTATG 1642
Db 131 GAACGACTCAAGACAGACAGAGAGAGCTGCGATATATGTTGTGACTTTGTATG 72
QY 1643 AAACATTTAGCTGATTTTGG 1662
Db 71 AAACATTTAGCTGATTTTGG 52

RESULT 12
AV816962/c 422 bp mRNA linear EST 01-APR-2002
LOCUS AV816962 RAFL9 Arabidopsis thaliana cDNA clone RAF109-92-K10 3',
DEFINITION mRNA sequence.
ACCESSION AV816962
VERSION AV816962.1 GI:19858828
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 422)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Saitou, M., Nakajima, M.,
Oono, Y., Sekurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.,
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki

Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060

Email: mseki@rkc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified lambda FIC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified Bluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES

source

1..422
Location/Qualifiers

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/clone="RAFL09-92-K10"

/clone_lib="RAFL9"

/dev_stage="plants at various developmental stages from germination to mature seeds"

/lab_host="DH10B"

/note="Site 1: BamHI; Site 2: SalI; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"

BASE COUNT

135 a 96 c 73 g 118 t

ORIGIN

Query Match

20.7%; Score 346.2; DB 10; Length 422;

Best Local Similarity 98.4%; Pred. No. 1.8e-96;

Matches 361; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

OY 1299 GACATGTGATGATTTTCAGGCTTAAGCTGTGACGATTCGATCAGACAAATTAATTCTTG 1358

DB 422 GACATGTGATGATTTTCAGGCTTAAGCTGTGACGATTCGATCAGACAAATTAATTCTTG 363

OY 1359 CAGTTACTGGGGATCTAGACAAGATGTTGCATGCAAGGTTATTTGAGCCCTATGCT 1418

DB 362 CAGTTACTGGGGATCTAGACAAGATGTTGCATGCAAGGTTATTTGAGCCCTATGCT 303

OY 1419 AATATGTGAGTTGCAAGAACCGGTGCTGCGATTGGCTCGGATCGGAGTGCATCTC 1478

DB 302 AATATGTGAGTTGCAAGAACCGGTGCTGCGATTGGCTCGGATCGGAGTGCATCTC 243

OY 1479 AAGTACCTTCGGATACCTCTTTTACAGGCTTAACCGTTGCAGAGTGCATCCAT 1538

DB 242 AAGTACCTTCGGATACCTCTTTTACAGGCTTAACCGTTGCAGAGTGCATCCAT 183

OY 1539 CGAATCATGAAACCTTTGGAAGGTAAAGTTTCATTACACAGTCTATGAAC--CTCAAA 1595

DB 182 CGAATCATGAAACCTTTGGAAGGTAAAGTTTCATTACACAGTCTATGAACCAACGCAAA 123

OY 1596 GACACAGAGAGAGCTGCTCATATATGTTTGACATTGTTTATGAACAATTAGCTG 1655

DB 122 GACACAGAGAGAGCTGCTCATATATGTTTGACATTGTTTATGAACAATTAGCTG 63

OY 1656 ATTTTGG 1662

DB 62 ATTTTGG 56

RESULT 13

AV805399/c

LOCUS AV805399 RAFL9 Arabidopsis thaliana CDNA clone RAFL09-42-M16 3',

DEFINITION mRNA sequence.

ACCESSION AV805399

VERSION AV805399.1

KEYWORDS EST.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE

AUTHORS

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 410)

Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,

Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,

Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.

and Shinozaki, K.

Large scale analysis of Arabidopsis full-length cDNA (2002b)

Unpublished (2002)

CONTACT: Motoaki Seki

Plant Functional Genomics Research Group

RIKEN Genomic Sciences Center

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-4359

Fax: 81-298-36-9060

Email: mseki@rkc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially

as reported previously (Seki et al., 1998). cDNA cleaved with BamHI

and XhoI was ligated to modified lambda FIC-1 vector (Carninci et

al., submitted for publication) digested with BamHI and SalI. This

clone is in a modified Bluescript vector. Please visit our web

site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further

details.

FEATURES

source

1..410
Location/Qualifiers

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/clone="RAFL09-42-M16"

/clone_lib="RAFL9"

/dev_stage="plants at various developmental stages from germination to mature seeds"

/lab_host="DH10B"

/note="Site 1: BamHI; Site 2: SalI; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"

BASE COUNT 129 a 94 c 74 g 113 t

ORIGIN

Query Match

20.2%; Score 338.2; DB 10; Length 410;

Best Local Similarity 98.1%; Pred. No. 5.5e-94;

Matches 354; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

OY 1306 CTAGATATTTTCAGGGCTAAGCTGTTGACGATTCGATCAGACAAATTAATTCTTGACAGCTTA 1365

DB 410 CTAGATATTTTCAGGGCTAAGCTGTTGACGATTCGATCAGACAAATTAATTCTTGACAGCTTA 351

OY 1366 CTGGGATTTAACAAGATGTTGCATGCAAGGTTATTTGAGCCCTATGCTATATGCTG 1425

DB 350 CTGGGATTTAACAAGATGTTGCATGCAAGGTTATTTGAGCCCTATGCTATATGCTG 291

OY 1426 AGCTTGCAAGAACCGGTGCTGCGATTGGCTCGGAATCGGAGTGCATCCAGGTACC 1485

DB 290 AGCTTGCAAGAACCGGTGCTGCGATTGGCTCGGAATCGGAGTGCATCCAGGTACC 231

OY 1486 TTCGTGATACCTCTTCTTTTACAGGCTTAACCGTTGCAGAGTGCATCCATGCAACAT 1545

DB 230 TTCGTGATACCTCTTCTTCTTTTACAGGCTTAACCGTTGCAGAGTGCATCCATGCAACAT 171

OY 1546 CAGAACTTTGAAGTTAAAGTTTCATTACACAGTCTATGAAC--CTCAAGACAGA 1601

DB 170 CAGAACTTTGAAGTTAAAGTTTCATTACACAGTCTATGAACCCCAAGTCAGAAAGCAGA 111

OY 1602 CAGAGACTGCGTGCATATATGTTTGATCTTTGTTTGAACAATTAAGCTGATTTTG 1661

DB 110 CAGAGACTGCGTGCATATATGTTTGATCTTTGTTTGAACAATTAAGCTGATTTTG 51

OY 1662 G 1662

DB 50 G 50

RESULT 14

AW310049

LOCUS AW310049 607 bp mRNA linear EST 02-DEC-2001

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccu@resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 411.
Location/Qualifiers
1. .675
source

BASE COUNT	193 a	125 c	170 g	187 t
ORIGIN				

Query Match	19.9%;	Score 333.2;	DB 10;	Length 675;
Best Local Similarity	70.3%;	Pred. NO. 2.8e-92;		
Matches 475; Conservative	0;	Mismatches 198;	Indels 3;	Gaps 2;

OY	625	TAGATATATAGCGGAACATGCACTTATCATATGAGGATACCTGAGATCCGTGGAAAAATGATGTG	684
Db	1	TGGATATCTCGCAACATTCGGGTACATTTGAGTCTACTGGAATCCAGGAGAGATGCGTG	60
OY	685	CTGTAGAAAGAAATTTGAAAAAGTTTCAGATCAGAGAGATTTGAAGGACAGGAAAGATAG	744
Db	61	CGGTTCAAAAGAAATTTCCAGCAAGTTGGAAATTTAAAGATATGCCAGAACTGGAAAG	-TTG 119
OY	745	CACGTGAAAGGAAAAAGATGGGTGCAAACGTCCCATTTTGGCGATTTTCAGACGATCT	804
Db	120	CATTAAAGAGGAAAAAGATGGGTGCATCTGCTCCATTTTGGGGATTCAGCTGCTCTTT	179
OY	805	ATTCAGATCTCAAGGAGACAGCCCTGTTAGTGT--TTCTCCAGATAGCAAAAAAGAGAC	862
Db	180	ATCCAGATCTTGAAGGAAAGAACACCTGTGATGCCATGCTGGGACGAAAAAATATGAAAC	239
OY	863	CATTGTCCTCCITCAAAAGCAACATCAGACAGGGGAGATGTTTATCCCGTTAGCCCATTTT	922
Db	240	CTGTGGCCCAACCTTGATTCACACTTGTGGGGGGAGATTTGTTATCCATATAGACCACCA	299
OY	923	TGACCCCAAGGATACATGCTATTTCTGCGACGCTGACGTGGGACCTTTCACAGCAGGAATAC	982
Db	300	TGGTTTTTACAGTCAATCAAGTACTTGTGATGCTCTCTGCTGGGTGTCCTCAATGAAAGAGATAC	359
OY	983	GAGTGGACTACGGTGCAGACTACTATCATTTGCTTGTGAAATGATATATTCACAGAGTTCTTAA	1042
Db	360	TAGTGGAAATTCGATACACACTCTATCCAGTCTTGGACAGATGCTCGGAGTTCTTAA	419
OY	1043	TATGTGACTGTGTGTTTTCCCTCGAAGGGGATACATATCCAGAGCTTGGCCGTAGGACA	1102
Db	420	CATTGTTTACAGGATATTTCTCTGAAGAGGCTATTAACATTCAGAGATTTAGCTGTAGGACA	479

OY	1103	TGCGAAGCAAGGCGATTTACAGCAATTACACAGTTAACTGCAACAGATGATCGCT	1162
Db	480	TGCGAAGTTGGAATGACTCTCTCGACTTACGACTGGTTCCTGCGACAGATGACTCAAT	539
OY	1163	CAGCAAAATGGTGCAAGCAACATTACAAACTCGATAGTGTGCATGAGTCCATGATCTTAC	1222
Db	540	AAGCAATTTGGTGCAAGCAACTATATATCAATAGAAAGCTACATGAGTGTGCGATATCAC	599
OY	1223	TCATTTGCCATTTTTCGAAAGAGAACTGATGCTGAATTAAGATTGCCGTGAACGCTGCTGC	1282
Db	600	CCACTTGCACATTTGCTGTGACGATGATTGCTGATAAAGATTCTTGTGAATGCTGCTGT	659
OY	1283	TAGAAGATGTCTCTG	1298
Db	660	ACCGCGTGAATGTACTG	675

```
Search completed: March 9, 2003, 21:40:21
Job time : 2579 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2003, 21:16:33 ; Search time 138 Seconds
(without alignments)
8097.319 Million cell updates/sec

Title: US-09-997-900-1
Perfect score: 1673
Sequence: 1 gctcttcacgtagcaaaaaa.....tgatttggcatttcg 1673

Scoring table:
Gapop 10.0 , Gapext 1.0

Searched: 478924 seqs, 33395956 residues
Total number of hits satisfying chosen parameters: 957848

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PC1_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PC10_NEW_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1673	100.0	1673	10	US-09-997-900-1
2	875.6	52.3	960	10	US-09-770-445-311
3	701.8	41.9	1721	10	US-09-732-618-17
4	665.8	39.8	1861	10	US-09-732-618-23
5	593.8	35.5	1435	10	US-09-732-618-15
6	590	33.3	1813	10	US-09-732-618-9
7	537.6	32.1	1297	10	US-09-732-618-5
8	419	25.0	4895	10	US-09-997-900-3
9	243.4	14.5	611	10	US-09-732-618-1
10	201.8	12.1	724	10	US-09-732-618-21
11	197	11.8	470	10	US-09-732-618-3
12	169.4	10.1	515	10	US-09-732-618-11
13	160.8	9.6	484	10	US-09-732-618-13
14	149.4	8.9	417	10	US-09-732-618-19
15	116.8	7.0	233	10	US-09-923-618-19
16	97.8	5.8	547	10	US-09-732-618-7
17	90	5.4	501	10	US-09-815-242-9274
18	90	5.4	501	10	US-09-815-242-9497
19	74.4	4.4	516	9	US-09-738-626-3458

C	20	64.8	3.9	640681	10	US-09-790-988-1	Sequence 1, Appl1
	21	64.6	3.9	522	10	US-09-815-242-9920	Sequence 9920, Ap
	22	62.6	3.7	489	10	US-09-815-242-3987	Sequence 3987, Ap
	23	62.6	3.7	522	10	US-09-815-242-7645	Sequence 7645, Ap
	24	56.8	3.4	492	10	US-09-815-242-7946	Sequence 7946, Ap
	25	55.4	3.3	492	10	US-09-761-782-1	Sequence 1, Appl1
	26	55.4	3.3	492	10	US-09-815-242-5926	Sequence 5926, Ap
	27	54.6	3.3	492	10	US-09-815-242-7159	Sequence 7159, Ap
	28	51.4	3.1	2000	9	US-09-938-8424-5198	Sequence 5198, Ap
C	29	38.4	2.3	8921	10	US-09-070-927A-68	Sequence 68, Appl
	30	36.6	2.2	273	10	US-09-878-574-14178	Sequence 14178, A
	31	36.2	2.2	2088	10	US-09-809-790-3	Sequence 3, Appl1
C	32	36.2	2.2	2088	10	US-09-809-617-3	Sequence 3, Appl1
C	33	35.8	2.1	1754	10	US-09-998-598-359	Sequence 39, Appl
C	34	35.8	2.1	2443	9	US-09-802-640-39	Sequence 268, Ap
C	35	35.2	2.1	3294	10	US-09-864-864-268	Sequence 550, Ap
	36	35.2	2.1	3344	10	US-09-964-824A-550	Sequence 2146, Ap
	37	35.2	2.1	3344	10	US-09-954-456-2146	Sequence 2260, Ap
	38	35.2	2.1	3344	10	US-09-880-107-2260	Sequence 1077, Ap
C	39	35	2.1	118	10	US-09-815-242-1077	Sequence 10, Appl
	40	35	2.1	2780	10	US-09-841-786-10	Sequence 8, Appl
	41	35	2.1	9726	10	US-09-841-786-8	Sequence 15, Appl
	42	35	2.1	11130	10	US-09-841-786-15	Sequence 112, Ap
C	43	34.4	2.1	3660	10	US-09-070-927A-112	Sequence 1370, Ap
	44	33.8	2.0	255	10	US-09-923-876-1370	Sequence 3620, Ap
	45	33.8	2.0	4140	10	US-09-764-877-3620	

ALIGNMENTS

RESULT 1
US-09-997-900-1
Sequence 1, Application US/09997900
Patent No. US20020053096A1
GENERAL INFORMATION:
APPLICANT: Kakefuda, Genichi
APPLICANT: Costello, Colleen
APPLICANT: Sun, Ming
APPLICANT: Hu, Weiming
TITLE OF INVENTION: Genes and Vectors for Confering Herbicide Resistance
TITLE OF INVENTION: In Plants
FILE REFERENCE: 043753/241148 (5849-20A)
CURRENT APPLICATION NUMBER: US/09/997,900
CURRENT FILING DATE: 2001-11-30
PRIOR APPLICATION NUMBER: 60/106,239
PRIOR FILING DATE: 1998-10-29
PRIOR APPLICATION NUMBER: 09/426,568
PRIOR FILING DATE: 1999-10-22
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 1673
TYPE: DNA
ORGANISM: Arabidopsis sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (42)..(1514)
OTHER INFORMATION: Mature Peptide
US-09-997-900-1

Query Match 100.0%: Score 1673; DB 10; Length 1673;
Best Local Similarity 100.0%: Pred. No. 0;
Matches 1673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTTCAGTAGCAAAAACCTGCGCTGCTGCAATGGCGCCATTCGTAA 60
|||||
DB 1 GTCCTTCAGTAGCAAAAACCTGCGCTGCTGCAATGGCGCCATTCGTAA 60
|||||
QY 61 GTCCTTCACCATCATTCGCTGCTGCTGAGATGGCATGTTCTTCTCTGCTTG 120
|||||
DB 61 GTCCTTCACCATCATTCGCTGCTGCTGAGATGGCATGTTCTTCTCTGCTTG 120
|||||

OY	121	TATCTCGACGCGGTATTCGCTTCCCGGCGAAGATTTCATATCTCTCCGGTATATCTCCG	180
OY	121	TATCTCGACGCGGTATTCGCTTCCCGGCGAAGATTTCATATCTCTCCGGTATATCTCCG	180
Db	121	TATCTCGACGCGGTATTCGCTTCCCGGCGAAGATTTCATATCTCTCCGGTATATCTCCG	180
OY	181	ACCGTGGCGATGAATATGGCTTAAGAGAATGGAAAGATCTCTTAAGAACGCTCATGGAGA	240
Db	181	ACCGTGGCGATGAATATGGCTTAAGAGAATGGAAAGATCTCTTAAGAACGCTCATGGAGA	240
OY	241	TCTCTGATGCGTCTTTCTCCGAAGCTTCATCTGCACTCCAAATTCGAAGGTGAGAGC	300
Db	241	TCTCTGATGCGTCTTTCTCCGAAGCTTCATCTGCACTCCAAATTCGAAGGTGAGAGC	300
OY	301	ACACAATTTGCTATTTTGTTGGAGAGAAAGCGAATGATTAATAGAAATTCGAGAGATG	360
Db	301	ACACAATTTGCTATTTTGTTGGAGAGAAAGCGAATGATTAATAGAAATTCGAGAGATG	360
OY	361	TTTGCAAGAGAGCATATACAATATTGAGAGTCTTGCTTGCTGTAACAGACAGCATC	420
Db	361	TTTGCAAGAGAGCATATACAATATTGAGAGTCTTGCTTGCTGTAACAGACAGCATC	420
OY	421	TATTCACCATATGTTGTCTGGAACCTGAAGGGTACTTCAGCAGGTCAATCGACAATCC	480
Db	421	TATTCACCATATGTTGTCTGGAACCTGAAGGGTACTTCAGCAGGTCAATCGACAATCC	480
OY	481	AGAAAGCTGCTTAATGTCTTAAGGTTGAAGATATCTCAAGTGAAGCCGCAAGTGGAGCTG	540
Db	481	AGAAAGCTGCTTAATGTCTTAAGGTTGAAGATATCTCAAGTGAAGCCGCAAGTGGAGCTG	540
OY	541	AGCTGATGCTGTGTAAAGTGAATGCATTCAGAAATCCAGGGCAGAGATCATGTGGCTAG	600
Db	541	AGCTGATGCTGTGTAAAGTGAATGCATTCAGAAATCCAGGGCAGAGATCATGTGGCTAG	600
OY	601	TTGACACATTCAGAGCAAGATTGTAGATATAGCGGACATGCATTCATCGAGATTA	660
Db	601	TTGACACATTCAGAGCAAGATTGTAGATATAGCGGACATGCATTCATCGAGATTA	660
OY	661	CTGAGATGCTCTGGAAAAATGATGCTGTGAAGAAATTTGAAAAAGTTTCAGATCAGAG	720
Db	661	CTGAGATGCTCTGGAAAAATGATGCTGTGAAGAAATTTGAAAAAGTTTCAGATCAGAG	720
OY	721	AGATTGTAAAGACAGGAAMATATAGCACTGAGAAAGGAAGAATGGCTCAACTGCTCCAT	780
Db	721	AGATTGTAAAGACAGGAAMATATAGCACTGAGAAAGGAAGAATGGCTCAACTGCTCCAT	780
OY	781	TTTGGCGATTTTCAGCAGCATCTTATCCAGATCTCAAGGAGCAAGCGCTGTAACTGTC	840
Db	781	TTTGGCGATTTTCAGCAGCATCTTATCCAGATCTCAAGGAGCAAGCGCTGTAACTGTC	840
OY	841	TTTCGATAGTCAAAAAAAGGAGCCATTGTCCTCAAAAGAAACATCAGCAGGGGAGATG	900
Db	841	TTTCGATAGTCAAAAAAAGGAGCCATTGTCCTCAAAAGAAACATCAGCAGGGGAGATG	900
OY	901	TTTATCCGTTGAGCCATTTTGTGACCCCAAGGTACATGTATTCGACGCTCATCTGG	960
Db	901	TTTATCCGTTGAGCCATTTTGTGACCCCAAGGTACATGTATTCGACGCTCATCTGG	960
OY	961	GACTTCTCAGTCAAGAGATACGAGTGAAGTACGAGTCACTACTCATTTGCTTTGAA	1020
Db	961	GACTTCTCAGTCAAGAGATACGAGTGAAGTACGAGTCACTACTCATTTGCTTTGAA	1020
OY	1021	ATGATATTCACGAGATCTTAATATTGTGACTGTGTTTCCCTCGAAGGGGATACATA	1080
Db	1021	ATGATATTCACGAGATCTTAATATTGTGACTGTGTTTCCCTCGAAGGGGATACATA	1080
OY	1081	TTCCAGAGCTTGGCCGTGAGCATGCTGAACCAAGGGCATTTTCACGCATTAACAAGTTA	1140
Db	1081	TTCCAGAGCTTGGCCGTGAGCATGCTGAACCAAGGGCATTTTCACGCATTAACAAGTTA	1140
OY	1141	TATCTGCAACAGATGATTCGCTCAGCAAAATTTGGTCAGCAACTTTTACAACCTCGATAGT	1200
Db	1141	TATCTGCAACAGATGATTCGCTCAGCAAAATTTGGTCAGCAACTTTTACAACCTCGATAGT	1200
OY	1201	TGCATGAGGCTCATGATCTTACTCATTTTGCCATTTTTCGAAAAGAACTGATGCTGATTA	1260

Db	1201	TCGATGAGGTCATGATCTTACTCATTTTGGCATTTTTCTAAGAGAGACTGATGCTGATTA	1260
Qy	1261	AGATTGGCGTCGAACGCGTGTGCTAGAGAAGATGTCCTGACATTTGCTAGTATTTTCACAGG	1320
Db	1261	AGATTGGCGTGAACGCGTGTGCTAGAGAAGATGTCCTGACATTTGCTAGTATTTTCACAGG	1320
Qy	1321	CTAAGCTGTTGACGATATCTGATCACACATTTACTTTGACGCTTACTGGGATCTTGACA	1380
Db	1321	CTAAGCGTTGACGATATCTGATCACACATTTACTTTGACGCTTACTGGGATCTTGACA	1380
Qy	1381	AGATGTTGACATCGCAAGGTTATTTGGAGCCCTATGCGTATATGTGAGGTTGCAGCAACCG	1440
Db	1381	AGATGTTGACATCGCAAGGTTATTTGGAGCCCTATGCGTATATGTGAGGTTGCAGCAACCG	1440
Qy	1441	GTCGTGGGCAATTGGGCTCGGATTCGGAGATGGACTCCCAAGTACCTGGTGATCTCCT	1500
Db	1441	GTCGTGGGCAATTGGGCTCGGATTCGGAGATGGACTCCCAAGTACTTGTGTGATCTCCT	1500
Qy	1501	TTCTTTTAACAGGCTAAACCGGTTGCAGAGTGCATCCATCGAACATCAGAACTTTGGAG	1560
Db	1501	TTCTTTTAACAGGCTAAACCGGTTGCAGAGTGCATCCATCGAACATCAGAACTTTGGAG	1560
Qy	1561	GTAAGAGTTTCATTACACAGCTATATGAACCTCAAGACAGACAGAGAGACTGCGTGATA	1620
Db	1561	GTAAGAGTTTCATTACACAGCTATATGAACCTCAAGACAGACAGAGAGACTGCGTGATA	1620
Qy	1621	TATGTTTGTGACTTGTCTTTATGAACAATTTAGTGATTTTGGGCTTCATTTG	1673
Db	1621	TATGTTTGTGACTTGTCTTTATGAACAATTTAGTGATTTTGGGCTTCATTTG	1673

```

RESULT 2
US-09-770-445-311/c
; Sequence 311, Application US/09770445
; Patent No. US2002023281A1
; GENERAL INFORMATION:
; APPLICANT: Goriach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Ralnes, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hufman, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 311
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-311

Query Match      52.3%: Score 875.6; DB 10; Length 960.
Best Local Similarity 98.7%: PId. No. 2.4e-262;

```


QY	285	TCGAAGGTGAGCAGCACACAAATTTTCAGTATTGTTGGACACGAAACCGGATGATTAAT	3444
Db	223	TCGGTTATGAAACGTACACACGCTATTCAGTTTGTGTGGTGAAGATGGGATGATCAAT	2822
QY	345	AGCAATTCGACGAGGTGTTTGCAGAAGAGATACAAATATTGAGCTCTTGCCTTGTCCTG	404
Db	283	CGAATTTGCTGGGGTTTTTGTGTAGAAGAGATTAACATCGAGTCAATTGGCTGGGGTTG	3424
QY	405	AACACAGCAAGAGCTCTATTACCACTAATTCTGTGTGAACATGAAAGGTTACTTCACAG	464
Db	343	AACAGAGGTTAAAGCTTTTATTACATATATTAGTGTGAGAAACAGCAAAATTTTAAACAG	4020

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 09:09:40 ; Search time 17 Seconds
(without alignments)
1217.974 Million cell updates/sec

Title: US-09-997-900-2

Perfect score: 2444
Sequence: 1 MAATSVSSSPSIRCLRSACS.....ARESGVDSKYLKRGYFLTLTG 491

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEM_PUB pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCR_NEM_PUB pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEM_PUB pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEM_PUB pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCRUS_PUBCOMB pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEM_PUB pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEM_PUB pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEM_PUB pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2444	100.0	491	US-09-997-900-2	Sequence 2, Appli
2	1692	69.2	476	US-09-732-618-18	Sequence 18, Appl
3	1493.5	61.1	483	US-09-732-618-10	Sequence 10, Appl
4	1476	60.4	449	US-09-732-618-24	Sequence 24, Appl
5	1476	60.4	449	US-09-732-618-25	Sequence 25, Appl
6	1395	57.1	365	US-09-732-618-16	Sequence 16, Appl
7	1302	53.3	350	US-09-732-618-6	Sequence 6, Appli
8	522	21.4	203	US-09-732-618-22	Sequence 22, Appl
9	499	20.4	133	US-09-732-618-22	Sequence 22, Appl
10	470.5	19.3	156	US-09-732-618-4	Sequence 4, Appli
11	436	15.2	156	US-09-732-618-12	Sequence 12, Appl
12	372.5	15.2	144	US-09-732-618-14	Sequence 14, Appl
13	329.5	13.5	135	US-09-732-618-20	Sequence 20, Appl
14	307	12.6	172	US-09-738-626-6958	Sequence 6958, Ap
15	305	12.5	166	US-09-815-242-13372	Sequence 13372, A
16	305	12.5	166	US-09-815-242-13595	Sequence 13595, A
17	277.5	11.4	163	US-09-761-782-2	Sequence 2, Appli
18	277	11.3	163	US-09-815-242-12044	Sequence 12044, A
19	276.5	11.3	163	US-09-815-242-10023	Sequence 10023, A

20	271.5	11.1	173	10	US-09-815-242-14018	Sequence 14018, A
21	269.5	11.0	173	10	US-09-815-242-11742	Sequence 11742, A
22	268.5	11.0	163	10	US-09-815-242-11256	Sequence 11256, A
23	267.5	10.9	163	10	US-09-815-242-5042	Sequence 5042, Ap
24	138	5.6	43	10	US-09-732-618-8	Sequence 8, Appli
25	135	5.5	828	9	US-09-738-626-5038	Sequence 5038, Ap
26	111	4.5	495	10	US-09-815-242-10315	Sequence 10315, A
27	107	4.4	500	10	US-09-815-242-13764	Sequence 13764, A
28	106.5	4.4	5215	9	US-09-860-846-2	Sequence 2, Appli
29	106.5	4.4	5215	9	US-09-861-289-2	Sequence 2, Appli
30	105.5	4.3	503	10	US-09-815-242-11400	Sequence 11400, A
31	105.5	4.3	3503	9	US-10-108-605-237	Sequence 237, App
32	105	4.3	495	10	US-09-815-242-5040	Sequence 5040, Ap
33	105	4.2	998	10	US-09-815-242-11716	Sequence 11716, A
34	103.5	4.2	978	10	US-09-893-817-4	Sequence 4, Appli
35	103.5	4.2	978	10	US-09-893-817-8	Sequence 8, Appli
36	102.5	4.2	96	10	US-09-741-669-362	Sequence 362, App
37	102.5	4.2	737	10	US-09-815-242-13699	Sequence 13699, A
38	101	4.1	84	10	US-09-815-242-5915	Sequence 5915, Ap
39	101	4.1	84	10	US-09-815-242-12930	Sequence 12930, A
40	101	4.1	84	10	US-09-815-242-13161	Sequence 13161, A
41	101	4.1	449	10	US-09-815-242-12423	Sequence 12423, A
42	101	4.1	452	10	US-09-815-242-5585	Sequence 5585, Ap
43	101	4.1	810	9	US-09-738-626-5524	Sequence 5524, Ap
44	100.5	4.1	608	10	US-09-815-242-11397	Sequence 11397, A
45	99.5	4.1	493	10	US-09-815-242-12052	Sequence 12052, A

ALIGNMENTS

```
RESULT 1
US-09-997-900-2
; Sequence 2, Application US/09997900
; Patent No. US20020053098A1
; GENERAL INFORMATION:
; APPLICANT: Kakefuda, Genichi
; APPLICANT: Costello, Colleen
; APPLICANT: Sun, Ming
; APPLICANT: Hu, Weiming
; TITLE OF INVENTION: Genes and Vectors for Conferring Herbicide Resistance
; FILE REFERENCE: 043753/241148 (5849-20A)
; CURRENT APPLICATION NUMBER: US/09/997,900
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/106,239
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 09/426,568
; PRIOR FILING DATE: 1999-10-22
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2:
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
US-09-997-900-2
Query Match      100.0%  Score 2444:  DB 10:  Length 491:
Best Local Similarity 100.0%:  Pred. No. 4.6e-203:
Matches 491:  Conservative 0:  Mismatches 0:  Indels 0:  Gaps 0:
```

```
QY 1 MAATSVSSSPSIRCLRSACSDSSPALVSTRVSFPKISYLGISSHRDEGKRRGEV 60
DB 1 MAATSVSSSPSIRCLRSACSDSSPALVSTRVSFPKISYLGISSHRDEGKRRGEV 60
QY 61 RSVGDKISDASESEASATPDKSVRKHTISVFGDESGMINIAGVFARGYNIESLAVG 120
DB 61 RSVGDKISDASESEASATPDKSVRKHTISVFGDESGMINIAGVFARGYNIESLAVG 120
QY 121 LNRDKALFTIIVCGTRVLYQVYIEQLKLVNVLKEDISSPOVEBELMLVYNAPESR 180
DB 121 LNRDKALFTIIVCGTRVLYQVYIEQLKLVNVLKEDISSPOVEBELMLVYNAPESR 180
```

QY 181 AEIMLVDTFFRRVNDIAEHALTIEVTGDPGKMAVERNLKKFQIREIVTGKIALRREK 240
DB 181 AEIMLVDTFFRRVNDIAEHALTIEVTGDPGKMAVERNLKKFQIREIVTGKIALRREK 240
QY 241 MGATAPFMEFSAASYPDLKEQAPVSVLRSSKKGAIYPOKETAGGVYVVEPFDPKVR 300
DB 241 MGATAPFMEFSAASYPDLKEQAPVSVLRSSKKGAIYPOKETAGGVYVVEPFDPKVR 300
QY 301 IDAHGGLTDEDTSGLRSHSTLSLVNDIPGVNLITGVFARGVNIQSLAAGHAETKGI 360
DB 301 IDAHGGLTDEDTSGLRSHSTLSLVNDIPGVNLITGVFARGVNIQSLAAGHAETKGI 360
QY 361 SRTTIVIPATDESYSKLVQOLYKLVNDVHEVDLTHLPFSERELMLIKIAVNAARRDVLD 420
DB 361 SRTTIVIPATDESYSKLVQOLYKLVNDVHEVDLTHLPFSERELMLIKIAVNAARRDVLD 420
QY 421 IASIFPAKAVDSDHTITLTQGLDGMVALORLLEPYGICEVARTGRVALARESGVDSK 480
DB 421 IASIFPAKAVDSDHTITLTQGLDGMVALORLLEPYGICEVARTGRVALARESGVDSK 480
QY 481 YLRGYSFLLTG 491
DB 481 YLRGYSFLLTG 491

RESULT 2

US-09-732-618-18
; Sequence 18, Application US/09732618
; Patent No. US20010044939A1

GENERAL INFORMATION:

; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Omolayo O.
; TITLE OF INVENTION: Small Subunit of Plant Acetolactate Synthase
; FILE REFERENCE: BB1435 US NA
; CURRENT FILING DATE: 2000-12-08
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/174,437
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Glycine max

US-09-732-618-18

Query Match 69.2%; Score 1692; DB 10; Length 476;
Best Local Similarity 71.5%; Pred. No. 3,8e-138;

Matches 337; Conservative 57; Mismatches 69; Indels 8; Gaps 2;

QY 19 GDSPPALVSTTRVSPFPAKISYLSGSISSHRGDEMGRMGFVRSDGKISDASFSSASSA 78
DB 14 CS-PNPKPLFTKRPLESPKPPSISTPHTRADKLISVSAATSSSGPPSPS----- 65
QY 79 TRSKVRKHTISVFGDESGMINRIAGVARGYNIESLAVGLNRKALFTTVGCTERY 138
DB 66 PARSKVRHTISVFGDESGMINRIAGVARGYNIESLAVGLNEDRAFTTVSGTDV 125
QY 139 LQOVIEQOLKLVNVLKVEDISSEPOVERELMLVKYNAHPESAELMWLVDTTRARVNDIA 198
DB 126 LQOVIEQOLKLVNVLKVEDISSEPOVERELMLIKYNAHAEIKWLVDTTRAKIVDS 185
QY 199 EHALTIEVTGDPGKMAVERNLKKFQIREIVTGKIALRREKMGATAPFMRSAASYPL 258
DB 186 EHSVITEVTGDPGKMAVERNLKKFQIREIVTGKIALRREKMGATAPFMRSAASYPL 245
QY 259 KEOAPVSVLRSSKKGAIYPOKETAGGVYVVEPFDPKVRHILDAHMGGLTDEDTSGLR 318
DB 246 EGRTPVNALVAKNNKVPKALDTIPVGGDYVPIEPSDGFVNVGLDAHMGVLDNEDTSGIR 305
QY 319 SHTLSLVNDIPGVNLITGVFARGVNIQSLAAGHAETKGISRTTIVIPATDESYSKLV 378
DB 319 SHTLSLVNDIPGVNLITGVFARGVNIQSLAAGHAETKGISRTTIVIPATDESYSKLV 378

DB 306 SHTLSLVNDIPGVNLITGVFARGVNIQSLAAGHAEGSLRTTIVPGTDESISKLV 365
QY 379 QOLYKLVNDVHEVDLTHLPFSERELMLIKIAVNAARRDVLDIASIFRAKAVDSDHTIT 438
DB 366 QOLYKLVNDVHEVDLTHLPFSERELMLIKIAVNAARRDVLDIASIFRAKAVDSDHTIT 425
QY 439 IQLTGDGMVALORLLEPYGICEVARTGRVALARESGVDSKYLGRYSFL 489
DB 426 LELTGDGMVALORLLEPYGICEVARTGRVALARESGVDSKYLGRYSFL 476

RESULT 3

US-09-732-618-10
; Sequence 10, Application US/09732618
; Patent No. US20010044939A1

GENERAL INFORMATION:

; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Omolayo O.
; TITLE OF INVENTION: Small Subunit of Plant Acetolactate Synthase
; FILE REFERENCE: BB1435 US NA
; CURRENT FILING DATE: 2000-12-08
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/174,437
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Zea mays
US-09-732-618-10

Query Match 61.1%; Score 1493.5; DB 10; Length 483;
Best Local Similarity 63.5%; Pred. No. 5.4e-121;

Matches 308; Conservative 61; Mismatches 83; Indels 33; Gaps 7;

QY 13 RCLRSACSDSPALVSTTRVSPFPAKISYLSGSISSHRGDEMGRMGFVRSDGKISDASF 72
DB 24 RCGAGSRVDSRPL--TPAVGFTA-----GPR---ARV--AVTAASS 58
QY 73 SPAS-SATP-----KSKVRKHTISVFGDESGMINRIAGVARGYNIESLAVGLNRK 125
DB 59 SPATGVTPFPPSRNSRVK RHTLSVFGDESGMINRIAGVARGYNIESLAVGLNRK 118
QY 126 ALFTTVGCTERYLQOVIEQOLKLVNVLKVEDISSEPOVERELMLVKYNAHPESAELMW 185
DB 119 ALFTTVGCTERYLQOVIEQOLKLVNVLKVEDISSEPOVERELMLIKYNAHAEIKWLV 178
QY 186 LVDTFRARVNDIAEHALTIEVTGDPGKMAVERNLKKFQIREIVTGKIALRREKMGATA 245
DB 179 LVRIKAEVNDISDVTTLTIEVGDGKMAVIAIKTKSGIRIARTGKIALRREKMGATA 238
QY 246 PWRFSASASYPDLKEQAPVSVLRSSKKGAIYPOKETAGGVYVVEPFDPKVRHILDA 304
DB 233 PWRFSASASYPDLKEQAPVSVLRSSKKGAIYPOKETAGGVYVVEPFDPKVRHILDA 298
QY 305 HMGGLTDEDTSGLRSHSTLSLVNDIPGVNLITGVFARGVNIQSLAAGHAETKGISRT 364
DB 299 HMGVMTDGPTEGCHTSTLSLVNDIPGVNLITGVFARGVNIQSLAAGPAKEGTSRT 358
QY 365 TVIPATDESYSKLVQOLYKLVNDVHEVDLTHLPFSERELMLIKIAVNAARRDVLDIASI 424
DB 359 TVIPATDESYSKLVQOLYKLVNDVHEVDLTHLPFSERELMLIKIAVNAARRDVLDIASI 418
QY 425 FRAKAVDSDHTITLTQGLDGMVALORLLEPYGICEVARTGRVALARESGVDSKYLGR 484
DB 419 FRAKAVDSDHTITLTQGLDGMVALORLLEPYGICEVARTGRVALARESGVDSKYLGR 478
QY 485 YSFL 489
DB 479 YSFL 483


```
RESULT 4
US-09-732-618-24
; Sequence 24, Application US/09732618
; Patent No. US20010044939A1
; GENERAL INFORMATION:
; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Farnodu, Omolayo O.
; TITLE OF INVENTION: Small subunit of Plant Acetolactate Synthase
; FILE REFERENCE: B01435 US NA
; CURRENT APPLICATION NUMBER: US/09/732,618
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/174,437
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Nicotiana glumabaginitfolia
US-09-732-618-24
```

```
Query Match          60.4%; Score 1476; DB 10; Length 449;
Best Local Similarity 79.1%; Pred. No. 1.6e-119;
Matches 291; Conservative 35; Mismatches 42; Indels 0; Gaps 0;
```

```
QY 99 MINRIAGVARRGYNIESLAVGLNRKALFTYVCGTERVLOQVITQLOKLVNVLKVEDI 158
DB 37 MINRIAGVARRGYNIESLAVGLNRKALFTYVCGTERVLOQVITQLOKLVNVLKVEDI 96
QY 159 SSEPOVERELMLKVNAPESRAEIMLVDTFRARVVDIAEHALTEVGDGPKMAVER 218
DB 97 SKEPOVERELMLKISADPKYRAEVMVLDVFRAKIVDSDSLTEVGDGPKMAVOR 156
QY 219 NLKFOIREIVRTGKTALRREKMGATAPWRSASYPDLKEQAPVSVLRSSKGAIVPQ 278
DB 157 NLKFGIREIARTGKTALRREKMGESAPWRSASYPDLKEGMSAGTISRITKRTPNGE 216
QY 279 KETSAGGDVYPVEPFDPKVRHRLDAHMGVLNDEDTSGLRSHLSLVNDIPGVNLIVG 338
DB 217 SMSMAGDVPVETDNDNSGVLDHMGVLDNEDTSGLRSHLSLVNDIPGVNLIVG 276
QY 339 VFARRGYNIOSLAVGHAETKISRITTVIPATDESYSKLVQOLYKLVADHEVNDLTHLP 398
DB 277 VFARRGYNIOSLAVGHAETKISRITTVIPATDESYSKLVQOLYKLVADHEVNDLTHLP 336
QY 399 SERELMLIKIAVNAARVDLIDIASIFRAKAVDSHTTTLQITGDLKMNALQRLLEPY 458
DB 337 SERELMLIKIAVNAARVDLIDIASIFRAKAVDSHTTTLQITGDLKMNALQRLLEPY 396
QY 459 GICEVART 466
DB 397 GICEVART 404
```

```
RESULT 5
US-09-732-618-25
; Sequence 25, Application US/09732618
; Patent No. US20010044939A1
; GENERAL INFORMATION:
; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Farnodu, Omolayo O.
; TITLE OF INVENTION: Small subunit of Plant Acetolactate Synthase
; FILE REFERENCE: B01435 US NA
; CURRENT APPLICATION NUMBER: US/09/732,618
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/174,437
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 25
```

```
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Nicotiana glumabaginitfolia
US-09-732-618-25
```

```
Query Match          60.4%; Score 1476; DB 10; Length 449;
Best Local Similarity 79.1%; Pred. No. 1.6e-119;
Matches 291; Conservative 35; Mismatches 42; Indels 0; Gaps 0;
```

```
QY 99 MINRIAGVARRGYNIESLAVGLNRKALFTYVCGTERVLOQVITQLOKLVNVLKVEDI 158
DB 37 MINRIAGVARRGYNIESLAVGLNRKALFTYVCGTERVLOQVITQLOKLVNVLKVEDI 96
QY 159 SSEPOVERELMLKVNAPESRAEIMLVDTFRARVVDIAEHALTEVGDGPKMAVER 218
DB 97 SKEPOVERELMLKISADPKYRAEVMVLDVFRAKIVDSDSLTEVGDGPKMAVOR 156
QY 219 NLKFOIREIVRTGKTALRREKMGATAPWRSASYPDLKEQAPVSVLRSSKGAIVPQ 278
DB 157 NLKFGIREIARTGKTALRREKMGESAPWRSASYPDLKEGMSAGTISRITKRTPNGE 216
QY 279 KETSAGGDVYPVEPFDPKVRHRLDAHMGVLNDEDTSGLRSHLSLVNDIPGVNLIVG 338
DB 217 SMSMAGDVPVETDNDNSGVLDHMGVLDNEDTSGLRSHLSLVNDIPGVNLIVG 276
QY 339 VFARRGYNIOSLAVGHAETKISRITTVIPATDESYSKLVQOLYKLVADHEVNDLTHLP 398
DB 277 VFARRGYNIOSLAVGHAETKISRITTVIPATDESYSKLVQOLYKLVADHEVNDLTHLP 336
QY 399 SERELMLIKIAVNAARVDLIDIASIFRAKAVDSHTTTLQITGDLKMNALQRLLEPY 458
DB 337 SERELMLIKIAVNAARVDLIDIASIFRAKAVDSHTTTLQITGDLKMNALQRLLEPY 396
QY 459 GICEVART 466
DB 397 GICEVART 404
```

```
RESULT 6
US-09-732-618-16
; Sequence 16, Application US/09732618
; Patent No. US20010044939A1
; GENERAL INFORMATION:
; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Farnodu, Omolayo O.
; TITLE OF INVENTION: Small subunit of Plant Acetolactate Synthase
; FILE REFERENCE: B01435 US NA
; CURRENT APPLICATION NUMBER: US/09/732,618
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/174,437
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-732-618-16
```

```
Query Match          57.1%; Score 1395; DB 10; Length 365;
Best Local Similarity 75.6%; Pred. No. 1.1e-112;
Matches 276; Conservative 38; Mismatches 49; Indels 2; Gaps 1;
```

```
QY 127 LETTIVCGTERVLOQVITQLOKLVNVLKVEDISSEPOVERELMLKVNAPESRAEIMML 186
DB 1 MFTIVSGTDRVNLQVITQLOKLVNVLNVEDLSKEPOVERELMLKINVEPDQRPVAVL 60
QY 187 VDFERRAVVDIAEHALTIVTGDPGKMAVERLKFQIREIVRTGKTALRREKMGATAP 246
DB 61 VDFERRAVVDISHTLITVETGDPGKMAVERLKFQIREIVRTGKTALRREKMGATAP 120
QY 247 FWRFSASYPDLKEQAPVSVLRSSKGAIVP--QKETSAGGDVYPVEPFDPKVRHRLDA 304
```



```
Db      2 VGLNKDKAMFTIVSGCTDVLNQVLEQLNKLVLNVLEDLSKEPQVERELMIKINVEPD 61
;
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6958
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6958

RESULT 13
US-09-732-618-20
; Sequence 20, Application US/09732618
; Patent No. US20010044939A1
; GENERAL INFORMATION:
; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Omolayo O.
; TITLE OF INVENTION: Small Subunit of Plant Acetolactate Synthase
; FILE REFERENCE: B01435 US NA
; CURRENT APPLICATION NUMBER: US/09/732,618
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/1174,437
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Trifolium aestivum
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (89)
; NAME/KEY: UNSURE
; LOCATION: (92)
; NAME/KEY: UNSURE
; LOCATION: (115)
; NAME/KEY: UNSURE
; LOCATION: (123)
; US-09-732-618-20

Query Match      13.5%; Score 329.5; DB 10; Length 135;
Best Local Similarity 54.3%; Pred. No. 2.6e-21;
Matches 75; Conservative 21; Mismatches 33; Indels 9; Gaps 4;

Qy      108 ARGGYIESLAVGLNDKALFTIVCGTERVLAQVIEQLQKLVNLKVEDISSEPOVERE 167
;
; TITLE OF INVENTION: Identification of Essential Genes In
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13372
; LENGTH: 166
; TYPE: PRT

Db      61 LMIKLNVEPDQADVMEFANVFESKCYFFXEQANSGGNWELL---GKSAQAQGTXXR 117
;
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6958
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6958

RESULT 14
US-09-738-626-6958
; Sequence 6958, Application US/09738626
; Patent No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKOYO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
```

```
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6958
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6958

Query Match      12.6%; Score 307; DB 9; Length 172;
Best Local Similarity 41.7%; Pred. No. 3.3e-19;
Matches 68; Conservative 36; Mismatches 49; Indels 10; Gaps 2;

Qy      82 SKVRKHTISVEYDESGMINRIAGVFARRGYNIESLAV-----GLNRDKALFTIVCGT 135
;
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6958
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6958

Db      4 SDVTRHILSVLVQVDVGLISRVSGMFTRRFNLVLSAKTETHGINR---ITVVVDAD 59
;
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6958
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6958

Qy      136 ERYLQOVIEQLQKLVNLKVEDISSEPOVERELMVKVNAHESRAELMVLVDTRARV 195
;
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6958
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6958

Db      60 ELNIEQITQKLNKLIPVLKVELEDETTIARAIMLVKVSADSTNRQIVDANIFRARRV 119
;
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6958
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6958

Qy      196 DIAEHALTIETVGDGPKMLAVERNLKKFOIREIVRTGKIALRR 238
;
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6958
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6958

Db      120 DVAPDSVIEISTGTGFKLALLDVMEPFGRILRLOSQTALNR 162
;
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6958
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6958

RESULT 15
US-09-815-242-13372
; Sequence 13372, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes In
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13372
; LENGTH: 166
; TYPE: PRT
```

; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13372

Query Match 12.5%; Score 305; DB 10; Length 166;
Best Local Similarity 39.4%; Pred No. 4, 7e-19;
Matches 67; Conservative 37; Mismatches 58; Indels 8; Gaps 3;

QY 306 WGLITDEDTSGLRSHLTSLLVNDIPGVNLIVGVFARGVNIOSLAVGHAETKGISRTT 365
||| | : : : ||| ||| :||| :||| :||| :|||
DB 4 WG-----CSSMR-RMLTFAKLQNRSGVNLNRFGTGVLRSQVNIESISVGATEDPNVSRTT 56
QY 366 VIP-ATDESVSKLVQOLKLVVHEVHDLTHLPFSERELMLKIAVNAARDVDLDIASI 424
:| : : : :||| :||| :||| :||| :||| :||| :||| :|||
DB 57 IIDVASHDEVEQIIKOLNROIDVIRIRIDITDKPHLEREVILVMSAPAEKRAEIIAIIOP 116
QY 425 FRAKAVDSDHRTTLOLTGDLDMVALORLLEPYGICEVARTGRVALARE 474
||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 117 FRATVVVAPSSITTIOMTGNAEKSEALNVRPYGIRNIARTGATGTFTRD 166

Search completed: March 10, 2003, 09:13:59
Job time : 18 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2003, 18:39:32 ; Search time 4658 Seconds

(without alignments)
10452.768 Million cell updates/sec

Title: US-09-997-900-1

Perfect score: 1673

Sequence: 1 gctcttcagtagcaaaaa.....tgatttggcctcattcg 1673

Scoring table: IDENTITY NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hgt_hum:*
31: em_hgt_inv:*
32: em_hgt_other:*
33: em_hgt_mus:*
34: em_hgt_pln:*
35: em_hgt_rtd:*
36: em_hgt_mam:*
37: em_hgt_vrt:*
38: em_sy:*
39: em_hgtgo_hum:*
40: em_hgtgo_mus:*
41: em_hgtgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1673	100.0	1673	6	AR194786	AR194786 Sequence
2	1644.6	98.3	1675	8	AY058881	AY058881 Arabidops
3	1474.4	88.1	1476	8	AY103307	AY103307 Arabidops
4	665.8	35.8	1860	8	NP1234901	AJ234901 Nicotiana
5	420	25.1	100655	8	AC006533	AC006533 Arabidops
6	419	25.0	4895	6	AR194787	AR194787 Sequence
7	156.6	9.4	6982	8	AF233069	AF233069 Galdieria
8	153.8	9.2	164921	8	AF022186	AF022186 Cyanidium
9	150.8	9.0	300450	1	AP005371	AP005371 Thermosyn
10	149.8	9.0	11318	1	AE013291	AE013291 Methanosa
11	147.8	8.8	191028	8	PP038804	U38804 Porphyra pu
12	147.2	8.8	7550	8	D63676	D63676 Cyanidium c
13	146	8.7	11759	1	AE011091	AE011091 Methanosa
14	145	8.7	121524	8	AF041468	AF041468 Guillardia
15	145	8.7	342850	1	AP003597	AP003597 Nostoc sp
16	137.4	8.2	8200	8	D63675	D63675 Cyanidiosch
17	132.2	7.9	105795	1	SYCCPNC	D64001 Synecocyst
18	130.6	7.8	116302	2	AC129715	AC129715 Medicago
19	128	7.7	82001	8	AB005242	AB005242 Arabidops
20	124.8	7.5	10481	1	AE012976	AE012976 Thermomana
21	123.8	7.4	104383	2	AP004227	AP004227 Oryza sat
22	123.8	7.4	138556	2	AP004881	AP004881 Oryza sat
23	118.4	7.1	13104	1	AE000985	AE000985 Archaeogl
24	111.4	6.7	12156	1	AE000759	AE000759 Aquifex a
25	104.8	6.3	15939	1	AE007128	AE007128 Mycobacte
26	104.8	6.3	70287	1	MTV012	AL021287 Mycobacte
27	104.4	6.2	10408	1	AE007812	AE007812 Clostridi
28	103.4	6.2	13072	1	U67473	U67473 Methanococc
29	98.4	5.9	44882	1	MLCB637	Z99263 Mycobacteri
30	98.4	5.9	348950	1	MLEPRTN7	AL583923 Mycobacte
31	97	5.8	2265	1	AF118061	AF118061 Methanoco
32	92	5.5	11643	1	AE000905	AE000905 Methanoba
33	91	5.4	525	6	AX100789	AX100789 Sequence
34	91	5.4	3291	6	AX100814	AX100814 Sequence
35	91	5.4	4281	1	BACILNB	L03181 Bacillus su
36	91	5.4	10426	6	AX100838	AX100838 Sequence
37	91	5.4	89009	1	BS275208	Z75208 B. subtilis
38	91	5.4	218410	1	BSUB0015	Z99118 Bacillus su
39	90.4	5.4	6211	6	BD003688	BD003688 Polynucle
40	90.4	5.4	10762	1	AE007355	AE007355 Streptoco
41	90.4	5.4	11274	1	AE008420	AE008420 Streptoco
42	90.4	5.4	11135	2	SPNEU1906	AL449928 Streptoco
43	89	5.3	4210	1	MSGILVB	L49392 Mycobacteri
44	87.4	5.2	3527	1	MAU35458	U35458 Methanococc
45	85.4	5.1	10029	1	AE012835	AE012835 Chlorobiu

ALIGNMENTS

RESULT 1
AR194786
LOCUS AR194786 1673 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6348643.
ACCESSION AR194786
VERSION AR194786.1 GI:20241378
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1673)
AUTHORS Kakefuda,G., Costello,C., Sun,M. and Hu,M.
TITLE DNA sequences encoding the arabinidopsis acetylhydroxy-acid synthase
small subunit and methods of use
JOURNAL Patent: US 6348643-A 1 19-FEB-2002;

FEATURES	Location/Qualifiers
SOURCE	1. 1673 /organism="unknown"
BASE COUNT	466 a 328 c 418 g 461 t
ORIGIN	
Query Match	100.0% Score 1673; DB 6; Length 1673;
Best Local Similarity	100.0% Pred. No. 0;
Matches 1673: Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 GTCCTTCACAGAGCAAAAAACCTTCGCGCTTCGTCGCAATGGCGGCATTTTCGTAA 60
Db	1 GCTCTTCACAGAGCAAAAAACCTTCGCGCTTCGTCGCAATGGCGGCATTTTCGTAA 60
OY	61 GTTCTTCACCATCTATTCGCTGCTTGAGATGCGCATGTCCTCGATCTTCCTGCTCTTG 120
Db	61 GTTCTTCACCATCTATTCGCTGCTTGAGATGCGCATGTCCTCGATCTTCCTGCTCTTG 120
OY	121 TATTCCTCAGCGGCTATTCGCTTCGCGGCAAAATTTTCATATCTTCGCTATATCTTGC 180
Db	121 TATTCCTCAGCGGCTATTCGCTTCGCGGCAAAATTTTCATATCTTCGCTATATCTTGC 180
OY	181 ACCGTGGCGATGAATGGGTAGAGATGGAAGATTCCTTGAAGCCGTCGATGGGAAGA 240
Db	181 ACCGTGGCGATGAATGGGTAGAGATGGAAGATTCCTTGAAGCCGTCGATGGGAAGA 240
OY	241 TCTGTATGTCGCTCTTCTTCGGAAGCTTCATCTGCACTCCAAATTCGAAGTGAGGAAGC 300
Db	241 TCTGTATGTCGCTCTTCTTCGGAAGCTTCATCTGCACTCCAAATTCGAAGTGAGGAAGC 300
OY	301 ACACAATTTCACTATTTGTTGGAGACGAAGCGGAATGATTAATAGATTTGCAGAGAGTGT 360
Db	301 ACACAATTTCACTATTTGTTGGAGACGAAGCGGAATGATTAATAGATTTGCAGAGAGTGT 360
OY	361 TTGCAAGGAGAGATATACAATATTGAGAGCTCTGCTGTTGCTGTAACAGAGACAAGGCTC 420
Db	361 TTGCAAGGAGAGATATACAATATTGAGAGCTCTGCTGTTGCTGTAACAGAGACAAGGCTC 420
OY	421 TATTCACCATAGTTGTCTGTGGAAGGCTTCTTCAGCAGGTCATCGACAATCC 480
Db	421 TATTCACCATAGTTGTCTGTGGAAGGCTTCTTCAGCAGGTCATCGACAATCC 480
OY	481 AGAAGCTGTTATGTTCTTAAAGGTTGAAGATATCTCAAGTGAAGCCGCAAGTGGAGGCTG 540
Db	481 AGAAGCTGTTATGTTCTTAAAGGTTGAAGATATCTCAAGTGAAGCCGCAAGTGGAGGCTG 540
OY	541 AGCTGATGCTTTTAAAGAGTATGCAATCCAGAAATCCAGGGCAGAGATCATGTGGCTAG 600
Db	541 AGCTGATGCTTTTAAAGAGTATGCAATCCAGAAATCCAGGGCAGAGATCATGTGGCTAG 600
OY	601 TTGACACATTCAGAGCAAGAGTTGATATATAGCGGAACATCATTTGACTATCGAAGTAA 660
Db	601 TTGACACATTCAGAGCAAGAGTTGATATATAGCGGAACATCATTTGACTATCGAAGTAA 660
OY	661 CTGGAGATTCCTGGAAAAATGATTTGCTGTAGAAAGAAATTTGAAAAAGTTTCAGATCAGAG 720
Db	661 CTGGAGATTCCTGGAAAAATGATTTGCTGTAGAAAGAAATTTGAAAAAGTTTCAGATCAGAG 720
OY	721 AGATTGTAAGACAGAGAAAGATAGCACTGAGAAAGGAAAGATGGTGCACATCGCTCCAT 780
Db	721 AGATTGTAAGACAGAGAAAGATAGCACTGAGAAAGGAAAGATGGTGCACATCGCTCCAT 780
OY	781 TTTGGCGATTTTTCAGCAGCATCTTATCCAGATCTCAAGGAGCAAGCGCTGTTAGTGTTC 840
Db	781 TTTGGCGATTTTTCAGCAGCATCTTATCCAGATCTCAAGGAGCAAGCGCTGTTAGTGTTC 840
OY	841 TTTCGAAGTAGCAAAAAAGAGCCATTGTCTCCCAAAAGCAAAATCATCAGCAGGGGAGATG 900
Db	841 TTTCGAAGTAGCAAAAAAGAGCCATTGTCTCCCAAAAGCAAAATCATCAGCAGGGGAGATG 900
OY	901 TTTATTCGCGTTAGGCAATTTTGTGACCCCAAGTATCATGTATTCGACGCTCACTGGG 960
Db	901 TTTATTCGCGTTAGGCAATTTTGTGACCCCAAGTATCATGTATTCGACGCTCACTGGG 960

QY	6961	GACCTCTCAGCAGACAGATACGAGTGGACATACAGCGTGCATACATCTATCATTCGCTGTAA	1020
Db	961	GACCTCTCAGCAGACAGATACGAGTGGACATACAGCGTGCATACATCTATCATTCGCTGTAA <th>1020</th>	1020
QY	1021	ATGATATATCCAGGAGTCTTATATATTGACCTGGTGTGTTTTCGCTCGAAGGCGATACATA <th>1080</th>	1080
Db	1021	ATGATATATCCAGGAGTCTTATATATTGACCTGGTGTGTTTTCGCTCGAAGGCGATACATA <th>1080</th>	1080
QY	1081	TCCAGAGCTTGGCCGTAGACATGCTGAACCAAGGCGATTTCACGCATTACACAGCTTA <th>1140</th>	1140
Db	1081	TCCAGAGCTTGGCCGTAGACATGCTGAACCAAGGCGATTTCACGCATTACACAGCTTA <th>1140</th>	1140
QY	1141	TACCGCAACAGATGAATCGGCTCAGCAAAATTGGTGACAGCACCCTTTACCAACTCGTAGTG <th>1200</th>	1200
Db	1141	TACCGCAACAGATGAATCGGCTCAGCAAAATTGGTGACAGCACCCTTTACCAACTCGTAGTG <th>1200</th>	1200
QY	1201	TGCATGAGTCCATGATCTTACTCATTTTGCATTTTTCGAAAGAGAACTGATGCTGATTA <th>1260</th>	1260
Db	1201	TGCATGAGTCCATGATCTTACTCATTTTGCATTTTTCGAAAGAGAACTGATGCTGATTA <th>1260</th>	1260
QY	1261	AGATTGCCGTAAACGCTGCTGCTAGAGAGATGTCCTGGACATTGCTATATTTTTCAGGG <th>1320</th>	1320
Db	1261	AGATTGCCGTAAACGCTGCTGCTAGAGAGATGTCCTGGACATTGCTATATTTTTCAGGG <th>1320</th>	1320
QY	1321	CTAAAGCTGTTGAGCTATCTGATCACAACAATTTACTTTGACCTTACAGGGGATCTAGACA <th>1380</th>	1380
Db	1321	CTAAAGCTGTTGAGCTATCTGATCACAACAATTTACTTTGACCTTACAGGGGATCTAGACA <th>1380</th>	1380
QY	1381	AGATGTTGCACTSCAAAGTTATTTAGAGCCCTATGATATATGTAGAGTTGCAAGAACCG <th>1440</th>	1440
Db	1381	AGATGTTGCACTSCAAAGTTATTTAGAGCCCTATGATATATGTAGAGTTGCAAGAACCG <th>1440</th>	1440
QY	1441	GTCGTGTGGCATTTGGCTCGTAATCGGGAGTGGACTCCAACTACCTTGTGTGATATCTCT <th>1500</th>	1500
Db	1441	GTCGTGTGGCATTTGGCTCGTAATCGGGAGTGGACTCCAACTACCTTGTGTGATATCTCT <th>1500</th>	1500
QY	1501	TTCTTTTACAGGCTAAACCGCTGACAGTGCATTCATCGAACAATCAGAACTTTGGAG <th>1560</th>	1560
Db	1501	TTCTTTTACAGGCTAAACCGCTGACAGTGCATTCATCGAACAATCAGAACTTTGGAG <th>1560</th>	1560
QY	1561	GTAAGAAGTTTCATTACACAGCTCATGAACCTCAAGAGACAGACAGACGCTCGATA <th>1620</th>	1620
Db	1561	GTAAGAAGTTTCATTACACAGCTCATGAACCTCAAGAGACAGACAGACGCTCGATA <th>1620</th>	1620
QY	1621	TATGTTGTGACTTGTGTTTATGAAGAACTTGAAGCTGATTTGGCGTCAATTCG <th>1673</th>	1673
Db	1621	TATGTTGTGACTTGTGTTTATGAAGAACTTGAAGCTGATTTGGCGTCAATTCG <th>1673</th>	1673
RESULT 2			
LOCUS	AY058881	1675 bp	mRNA linear PLN 03-NOV-2001
DEFINITION	Arabidopsis thaliana At2g31810/f20M17.15	mRNA, complete cds.	
ACCESSION	AY058881		
VERSION	AY058881.1	GI:16604522	
KEYWORDS	FLI CDNA.		
SOURCE	Arabidopsis thaliana.		
ORGANISM	Arabidopsis thaliana.		
REFERENCE			
AUTHORS	Shinozaki, K., Chen, H., Cheuk, R., Kim, C. J., Kosame, E., Meyers, M. C., Banh, J., Bowser, L., Carninci, P., Dale, J. M., Goldsmith, A. D., Hayashizaki, Y., Ishida, J., Jiang, P. X., Jones, T., Kamuya, A., Karlin-Neuman, G., Kawai, J., Lam, B., Lee, J. M., Lin, J., Liu, S. X., Miranda, M., Natusaka, M., Nguyen, M., Onodera, C. S., Palm, C. J., Pham, P. K., Quach, H. L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C. C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R. W., Theologis, A. and Ecker, J. R.		
TITLE	Arabidopsis cDNA clones		
JOURNAL	Unpublished		

REFERENCE 2 (bases 1 to 1675)
AUTHORS Shin, P., Chen, H., Cheuk, R., Kim, C.-J., Koesema, E., Meyers, M. C.,
Banh, J., Bowser, L., Carinici, P., Dale, J. M., Goldsmith, A. D.,
Hayashizaki, Y., Ishida, J., Jiang, P.-X., Jones, T., Kamiya, A.,
Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J. M., Lin, J., Liu, S. X.,
Miranda, M., Narusaka, M., Nguyen, M., Onodera, C. S., Palm, C. J.,
Pham, P. K., Quach, H. L., Sakurai, T., Satou, M., Seki, M., Southwick, A.,
Tang, C. C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,
Shiozaki, K., Davis, R. W., Theologis, A. and Ecker, J. R.
TITLE Direct Submission
JOURNAL Submitted (05-SEP-2001) Salk Institute Genomic Analysis Laboratory
(Signal), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA

FEATURES
source Location/Qualifiers
1. 1675
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="2"
/clone="RAF109-09-C24(R14179)"
/note="ecotype: Columbia"
50..1525
/note="putative acetolactate synthase"
/codon_start=1
/product="At2g31810/F20M17.15"
/protein_id="AAL24267.1"
/db_xref="GI:16604523"
/translation="MAAISVSSSPSRCLRSACSDSPALVSTPVSFPKAKISLSGI
SSHRDEMKRMEGVRSVSDKISDPSFSEASSATPKSVKRNHTTSVFQDGSIMNR
IAGVFAKRGYNIESLAVSLGNRKALFTTIVCGTEVNLQVIOQLKLVNLVLEIIS
EPQVRELMILVKNHPESRAEIMLVDFRFRVVDIEHALITEVTPGPMIAVER
NLKRFQIREIVYTGKIALRREKMGATAPFRFSAASYPDLKQAPSVYRSKGAIV
POKETASGDVYVPEPDPKVRHILDMGLTDMGTLSDGTSLSILVNDIPGLN
ITGVFAKRGYNIOISLAVGHAKTISRITTYTPATPDESVSLNOOLKLVNHYEVD
LTHLPFSEKELMLIKITAVNNAARRVDLIASIFRAKADVSSHITLQTLGDLDMKVA
LORLEPFGICVATGRVALRESGVDSKILRGVSFPLTG"

CDS
BASE COUNT 469 a 331 c 417 g 458 t
ORIGIN

Query Match 98.3%; Score 1644.6; DB 8; Length 1675;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

1 GTCCTTCAGTAGCAAAAAACCTCGGCTCGTCGTCATATGCGGCCATTCTGTAA 60
|||||
9 GTCCTTCAGTAGCAAAAAACCTCGGCTCGTCGTCATATGCGGCCATTCTGTAA 68
|||||
61 GTCCTTCAGTAGCAAAAAACCTCGGCTCGTCGTCATATGCGGCCATTCTGTAA 120
|||||
69 GTCCTTCAGTAGCAAAAAACCTCGGCTCGTCGTCATATGCGGCCATTCTGTAA 128
|||||
121 TATCCGCGACGCGTATGCTTCCCGCGCAAGATTTCATATCTCTCCGGATATCTTCCG 180
|||||
129 TATCCGCGACGCGTATGCTTCCCGCGCAAGATTTCATATCTCTCCGGATATCTTCCG 188
|||||
181 ACCGTGGCATGAATGGGTAAGAGATGGAAGATGCTTTAGTAAGCGTGCATGGGAGA 240
|||||
189 ACCGTGGCATGAATGGGTAAGAGATGGAAGATGCTTTAGTAAGCGTGCATGGGAGA 248
|||||
241 TCTCTGATGCGCTTTCTCCGAAGCTTCATCTGCACTCCAAAATCGAAGTGAAGAAC 300
|||||
249 TCTCTGATGCGCTTTCTCCGAAGCTTCATCTGCACTCCAAAATCGAAGTGAAGAAC 308
|||||
301 ACACATTTTCAGTATTTGTTGGAGAGAAAGGATGATTAATAGATTTGACGAGACTGT 360
|||||
309 ACACATTTTCAGTATTTGTTGGAGAGAAAGGATGATTAATAGATTTGACGAGACTGT 368
|||||
361 TTGCAAGAGAGAGATCAATATTGAGAGTCTTCTGTGGTCTGAACAGAGACAGGCTC 420
|||||
369 TTGCAAGAGAGAGATCAATATTGAGAGTCTTCTGTGGTCTGAACAGAGACAGGCTC 428
|||||
421 TATTCACATAGTGTCTGTGGAAGTGAAGGTTACTTCACGAGGTCATGACCAACTCC 480
|||||

Db 429 TATTCACATAGTGTCTGTGGAAGTGAAGGTTACTTCACGAGGTCATGACCAACTCC 488
481 AGAAGCTCGTAAATGTTCTAAAGGTGAAGATATTCAGAGTGAGCGCGCAAGTGACGCTG 540
|||||
489 AGAAGCTCGTAAATGTTCTAAAGGTGAAGATATTCAGAGTGAGCGCGCAAGTGACGCTG 548
|||||
541 AGCTGATGCTTGTAAAGTGAATGACATTCAGAAATTCAGAGGAGAGATCATGTGCTGTG 600
|||||
549 AGCTGATGCTTGTAAAGTGAATGACATTCAGAAATTCAGAGGAGAGATCATGTGCTGTG 608
|||||
601 TTGACACATTCAGAGCAAGAGTGTAGATATAGCGCAACATGCTATTCAGAGCTAA 660
|||||
609 TTGACACATTCAGAGCAAGAGTGTAGATATAGCGCAACATGCTATTCAGAGCTAA 668
|||||
661 CTGACATGCTGGAAGAAATGATTCGTGTGAAGAAATTTGAAAGTTTCAGATCAGAG 720
|||||
669 CTGACATGCTGGAAGAAATGATTCGTGTGAAGAAATTTGAAAGTTTCAGATCAGAG 728
|||||
721 AGATTTGAAGAGAGAGAAAGATAGACAGAGAGAGAGAAAGATGGTGCACACTGCTCAT 780
|||||
729 AGATTTGAAGAGAGAGAAAGATAGACAGAGAGAGAGAAAGATGGTGCACACTGCTCAT 788
|||||
781 TTGGCGGATTTTCAGCAGCATCTTATCCAGATCTCAAGAGCAAGCGCTGTAGTGTTC 840
|||||
789 TTGGCGGATTTTCAGCAGCATCTTATCCAGATCTCAAGAGCAAGCGCTGTAGTGTTC 848
|||||
841 TTGACAGTGAAGCAAAAAAGAGAGCATCTGCTCCCAAAAGAAACATGAGAGGGGGAGATG 900
|||||
849 TTGACAGTGAAGCAAAAAAGAGAGCATCTGCTCCCAAAAGAAACATGAGAGGGGGAGATG 908
|||||
901 TTATATCCCGTTGAGCCATTTTGGACCCCAAGATGATGTTATTCAGAGCTGCTGAG 960
|||||
909 TTATATCCCGTTGAGCCATTTTGGACCCCAAGATGATGTTATTCAGAGCTGCTGAG 968
|||||
961 GACTTCTCACTGACAGAGATAGAGTACGATGACGTCGATCTATCATATCTGTGTA 1020
|||||
969 GACTTCTCACTGACAGAGATAGAGTACGATGACGTCGATCTATCATATCTGTGTA 1028
|||||
1021 ATGATATTCAGAGAGTCTTAAATATGTCAGCTGTGTTTCCGTCCAGAGGGATTAATA 1080
|||||
1029 ATGATATTCAGAGAGTCTTAAATATGTCAGCTGTGTTTCCGTCCAGAGGGATTAATA 1088
|||||
1081 TCCAGAGCTTTGGCGGTAGAGCATGCTGAAGAACCAAGGCAATTCACCATTAACAAGTTA 1140
|||||
1089 TCCAGAGCTTTGGCGGTAGAGCATGCTGAAGAACCAAGGCAATTCACCATTAACAAGTTA 1148
|||||
1141 TACCTGCAACAGATGATGCTGACGCAAAATGGTGCAGCAACTTACAACTCGTAGATG 1200
|||||
1149 TACCTGCAACAGATGATGCTGACGCAAAATGGTGCAGCAACTTACAACTCGTAGATG 1208
|||||
1201 TGCATAGAGTCCATGATCTTACTCATTTTCTGTAAGAGAACTGATCTGATTA 1260
|||||
1209 TGCATAGAGTCCATGATCTTACTCATTTTCTGTAAGAGAACTGATCTGATTA 1268
|||||
1261 AGATTCGCGTGAAGCGTGTCTGAGAGAGATGCTCGGCACTTGTGATGATTTTACAGG 1320
|||||
1269 AGATTCGCGTGAAGCGTGTCTGAGAGAGATGCTCGGCACTTGTGATGATTTTACAGG 1328
|||||
1321 CTAAAGCTGTTGACGATCTGATCACAATTAATCTTGGACGTTACTGGGATCTGAGACA 1380
|||||
1329 CTAAAGCTGTTGACGATCTGATCACAATTAATCTTGGACGTTACTGGGATCTGAGACA 1388
|||||
1381 AGATGCTGACGTCGCAAGAGTTATGAGAGCCCTATGATGATGATGATTTGCAAGAACG 1440
|||||
1389 AGATGCTGACGTCGCAAGAGTTATGAGAGCCCTATGATGATGATGATTTGCAAGAACG 1448
|||||
1441 GTCGTGTGACATTTGCTGTAATTCGAGAGTGCATCCAAATCTTCTGATGATCTGCT 1500
|||||
1449 GTCGTGTGACATTTGCTGTAATTCGAGAGTGCATCCAAATCTTCTGATGATCTGCT 1508
|||||
1501 TTCTTTTAAAGAGCTTAACCGCTTGCAGAGTGCATTCATGCAATCAAGAACTTTGGAAG 1560
|||||
1509 TTCTTTTAAAGAGCTTAACCGCTTGCAGAGTGCATTCATGCAATCAAGAACTTTGGAAG 1568
|||||

OY	1561	GTAAGAGTTTCATCTACAGTCTATGAAAC---CTCAAGAAGACAGAGACTCGTG	1617
Db	1569	GTAAAGATTTCATCTACAGCTTATGAACCAACTAAGACACAGAGAGACTGGTAG	1628
OY	1618	ATAATGTTTTGACTTGTATTATGAACAATTAAGCTGATTTTGGC	1664
Db	1629	ATAATGTTTGTGCATTTGTTATATGAACAATTAAGCTGATTTTGGCC	1675
RESULT 3			
AY103307		1476 bp mRNA linear PLN 01-JUL-2002	
LOCUS			
DEFINITION	Aribidopsis thaliana Atcg31810/F2OM17.15 mRNA, complete cds.		
ACCESSION	AY103307		
VERSION	AY103307.1 GI:21655294		
KEYWORDS	FLI_CDNA.		
SOURCE	Thale cress.		
ORGANISM	Aribidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Aribidopsis.		
REFERENCE	1 (bases 1 to 1476) Kim,C.-J., Chen,H., Cheuk,R., Meyers,M.C., Shinn,P., Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Tortum,M., Wu,H.C., Yanada,K., Yamamura,Y., Yu,G., Yu,S., Shinzaki,K., Davis,R.W., Theologis,A. and Ecker,J.R. Arabidopsis ORF clones Unpublished		
TITLE	2 (bases 1 to 1476)		
JOURNAL	Kim,C.-J., Chen,H., Cheuk,R., Meyers,M.C., Shinn,P., Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Tortum,M., Wu,H.C., Yanada,K., Yamamura,Y., Yu,G., Yu,S., Shinzaki,K., Davis,R.W., Theologis,A. and Ecker,J.R. Direct Submission		
COMMENT	Submitted (11-MAY-2002) Salk Institute Genomic Analysis Laboratory (SIGNAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : "RIKEN Arabidopsis Full-Length cDNA") : Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinzaki,K.		
FEATURES	The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Kim,C.J., Chen,H., Cheuk,R., Meyers,M.C., Shinn,P., Banh,J., Bowser,L., Chang,E., Dale,J.M., Goldsmith,A.D., Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Southwick,A., Tang,C.C., Tortum,M., Wu,H.C., Yanada,K., Yamamura,Y., Yu,G., Yu,S., Davis,R.W., Theologis,A., and Ecker,J.R.		
SOURCE	Kim,C.J. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinzaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk) contributed equally to this work as PIs. Location/Qualifiers 1..1476 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="2" /clone="U14179" /note="This clone is in pUNI 51 ecotype: Columbia"		

CDS		1.. 1476	/note="putative acetolactate synthase"
		/codon_start=1	
		/product="Ac2g31810/F20M17.15"	
		/protein_id="AA06359.1"	
		/db_xref="GI:21655295"	
		/translation="MAAIVSSSPSIRCLRSGCDSDSPALVSSRVSFPAKISVLGSI SSHHDGMCKRMGEFVRSDGKIISDAFSFSSASSTPKSKYRKHILSYFDQEDMINR IAGFPARRGYNIESLAVGLNRDKALFLTIYVCGRVRLDQYLIEJOLKVNLVKEDISS EPOVERELMLTVKNAPHESEAETIMLVDTFRARVRLDEHALTLEVGDGKMIAVER NLKFQIREIVRGKIALREKMGATPEPRFSAASPDLKEQAPVLSRSKKGAIV PQKTSAGDVGYPEVEPFDPVHRIDAHNGLLEDDPTSLRSHTLSLVNDIPGVNI LVNGVFARRGNINOSLAVHAETRGISRITTYVPADVESIKYLOQLKYLDVEHVD LTHPFERSELMLKIVNAARRDVDIADISIFRAKVDVSDHITIQLTGDLDMVA LQRLPEREGICEVARTGRVALARSQSDYLRKYSPLTG"	
BASE COUNT	409 a	289 c	378 g 400 t
ORIGIN			
Query Match	88.1%	Score 1474.4:	DB 8; Length 1476;
Best Local Similarity	99.9%	Pred. No. 0;	
Matches 1475; Conservative	0; Mismatches	1; Indels	Gaps 0;
Db	42	ATGGCGGCCATTTCGTGAAGTTCTTCACCACTATATCCGTCCTTGAGATCGGCAGTTC	101
Db	1	ATGGCGGCCATTTCGTGAAGTTCTTCACCACTATATCCGTCCTTGAGATCGGCAGTTC	60
Qy	102	GATTCCTCCTCGCTCTTGTAATCCTCGACGCGGTATCGTTCCGGCGAGAATTCATAT	161
Db	61	GATTCCTCCTCGCTCTTGTAATCCTCGACGCGGTATCGTTCCGGCGAGAATTCATAT	120
Qy	162	CTTCGCCGATATATCTTCCGACCGGTGGCATGAATGGGTAAAGATAATGAGATTCGTT	221
Db	121	CTTCGCCGATATATCTTCCGACCGGTGGCATGAATGGGTAAAGATAATGAGATTCGTT	180
Qy	222	AGAACGTCGATGGGAATCTCTGATCGTCCTTCTCCGAAGCTTCATCTCGCACTCCA	281
Db	181	AGAACGTCGATGGGAATCTCTGATCGTCCTTCTCCGAAGCTTCATCTCGCACTCCA	240
Qy	282	AAATCGAAGGTGAGGAACACACAATTCACGATATTTGTTGAGACGAAGCGGAATGAT	341
Db	241	AAATCGAAGGTGAGGAACACACAATTCACGATATTTGTTGAGACGAAGCGGAATGAT	300
Qy	342	AATAGATTGAGGAGTGTTTGGCAAAGAGAGATACATATTGAGAGCTTGCTGTGGT	401
Db	301	AATAGATTGAGGAGTGTTTGGCAAAGAGAGATACATATTGAGAGCTTGCTGTGGT	360
Qy	402	CTGAACAGACACAAGGCTCTATTCAACCCTATGTCGTGTGAACGAAAGGACTTCAG	461
Db	361	CTGAACAGACACAAGGCTCTATTCAACCCTATGTCGTGTGAACGAAAGGACTTCAG	420
Qy	462	CAGTTCATCGAACCACTCCAGAAAGCTCGTTAATGTTCTAAAGGTTGAAGTATCTCAAGT	521
Db	421	CAGTTCATCGAACCACTCCAGAAAGCTCGTTAATGTTCTAAAGGTTGAAGTATCTCAAGT	480
Qy	522	GAGCCGCAAGTGGAGCGTGCATGCTTTTAAAAGTGAATGCACATCCGAATCCAGG	581
Db	481	GAGCCGCAAGTGGAGCGTGCATGCTTTTAAAAGTGAATGCACATCCGAATCCAGG	540
Qy	582	GCAGAGATCATGTGGCTATGTTGACCAATTCAGACCAAGATGTTAGATATATGCGGAACAT	641
Db	541	GCAGAGATCATGTGGCTATGTTGACCAATTCAGACCAAGATGTTAGATATATGCGGAACAT	600
Qy	642	GCATTGACTATCGAGGTAACTGAGATCTTGAAAAATGATTTGCTGTGAAAGAAATTTG	701
Db	601	GCATTGACTATCGAGGTAACTGAGATCTTGAAAAATGATTTGCTGTGAAAGAAATTTG	660
Qy	702	AAAAAGTTTCAGATCAGAGAGATTGTAAGACAGAGAAAGATAGCACTGAGAAGGAAAG	761
Db	661	AAAAAGTTTCAGATCAGAGAGATTGTAAGACAGAGAAAGATAGCACTGAGAAGGAAAG	720
Qy	762	ATGGGTGCAACTGCTCCATTTTGGCGATTTTCACGAGCATCTTATCCAGATCTCAAGAG	821
Db	721	ATGGGTGCAACTGCTCCATTTTGGCGATTTTCACGAGCATCTTATCCAGATCTCAAGAG	780

QY	822	CAAGGCGCTGTATAGTTCTTCGAAAGTAGCAAAAAAGAGCATTTGCCCAAAAGAA	881
Db	781	CAAGGCGCTGTATAGTTCTTCGAAAGTAGCAAAAAAGAGCATTTGCCCAAAAGAA	840
QY	882	ACATCAGCAGGAGGAGATGTTATTCGCCGTGTAGCCATTTTGTGACCCCAAGATACATGCT	941
Db	841	ACATCAGCAGGAGGAGATGTTATTCGCCGTGTAGCCATTTTGTGACCCCAAGATACATGCT	900
QY	942	ATTCTCGACGCTCACTGGGGACCTTCACATGACGAAGAATACAGATGACATACGGTGCAT	1001
Db	901	ATTCTCGACGCTCACTGGGGACCTTCACATGACGAAGAATACAGATGACATACGGTGCAT	960
QY	1002	ACTCATCATGCTGTAAATGATATTCACAGAGCTTTCTTAATTTGTGACGTGTGTTTC	1061
Db	961	ACTCATCATGCTGTAAATGATATTCACAGAGCTTTCTTAATTTGTGACGTGTGTTTC	1020
QY	1062	GCTCGAAGGGATACAAATATCCAGAGCTTGGCCGTAGACATGCTGAAACCAAGGCGATT	1121
Db	1021	GCTCGAAGGGATACAAATATCCAGAGCTTGGCCGTAGACATGCTGAAACCAAGGCGATT	1080
QY	1122	TCACGCAATTACACAGATTATACCTGCAACAGATGAATGCTCAGCAAAATTTGGTGACAA	1181
Db	1081	TCACGCAATTACACAGATTATACCTGCAACAGATGAATGCTCAGCAAAATTTGGTGACAA	1140
QY	1182	CTTTACAAACCTCGTAAGATGTCATGAGGTCCATGATCTTACTCATTTTCCATTTTCTGAA	1241
Db	1141	CTTTACAAACCTCGTAAGATGTCATGAGGTCCATGATCTTACTCATTTTCCATTTTCTGAA	1200
QY	1242	AGAACACGATGCGATTAATGAATGCGCGTGAACGCTGCTGTAGAAAGATGCTCTGGAC	1301
Db	1201	AGAACACGATGCGATTAATGAATGCGCGTGAACGCTGCTGTAGAAAGATGCTCTGGAC	1260
QY	1302	ATTGCTAGTATTTTCAGGGCTAAAGCTGTGTAGCGTATTCGATCAACAAATTTTTCGAG	1361
Db	1261	ATTGCTAGTATTTTCAGGGCTAAAGCTGTGTAGCGTATTCGATCAACAAATTTTTCGAG	1320
QY	1362	CTTACTGGGAGATTCAGCAACATGTTGCACTGCAACAGATTAATTCGACCCCTATGCTATTA	1421
Db	1321	CTTACTGGGAGATTCAGCAACATGTTGCACTGCAACAGATTAATTCGACCCCTATGCTATTA	1380
QY	1422	TGTAGAGTTGCAGAAACCGGTCGCTGTGGCATTTGGCTCTGTAATGGGAGTGGACTCCAAAG	1481
Db	1381	TGTAGAGTTGCAGAAACCGGTCGCTGTGGCATTTGGCTCTGTAATGGGAGTGGACTCCAAAG	1440
QY	1482	TACCTTGTGTGATCTCTCTTTCTTTAAACAGGCTTA	1517
Db	1441	TACCTTGTGTGATCTCTCTTTCTTTAAACAGGCTTA	1476

RESULT 4
NPL234901
 LOCUS
 DEFINITION
 NPL234901 1860 bp mRNA linear PLN 26-SEP-1999
 LOCUS
 Nicotiana plumbaginifolia mRNA for acetolactate synthase small subunit.
 AJ234901
 AJ234901.1 GI:5931760
 VERSION
 keywords
 SOURCE
 ORGANISM
 Nicotiana plumbaginifolia.
 Nicotiana plumbaginifolia
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; eunasterids I; Solanales; Solanaceae; Nicotiana.
 1 (bases 1 to 1860)
 Hershhey,H.P., Schwartz,L.J., Gale,J.P., and Abell,L.M.
 Cloning and functional expression of the small subunit of acetolactate synthase from Nicotiana plumbaginifolia
 Plant Mol. Biol. 40 (5), 795-806 (1999)
 JOURNAL
 PUBLISHED
 PUBMED
 10487214
 2 (bases 1 to 1860)
 Hershhey,H.P.
 REFERENCE
 AUTHORS
 Direct Submission

JOURNAL	Submitted (04-NOV-1998) Hershey H.P., Agricultural Products, E. I. du Pont de Nemours and Co., P.O. Box 80402, Wilmington, DE 19880-0402, USA
FEATURES	Location/Qualifiers
Source	1. 1860
CDS	/organism="Nicotiana glumbaginifolia" /db_xref="taxon:4092" 181..1530 /codon_start=1 /product="acetolactate synthase small subunit" /protein_id="CAB56614.1" /db_xref="GI:5931761" /translation="MEHIQIRLTLPSPDKRLAIRFKCLLYMKVEMINRIAGV FARGYNIESLAVGLNKDALEFTIVSGTERVLOOQMEOLKLVNIVKIDELSPQV ERELMLKISADPKYRAEVAMVINDVFAKTIYDSDLETSLEVDSPCKVAQNPISLK FGIREIARTKIDILIREKMGESAFPRFSAASYIDELGASAGTISRTIKTPGESK SMEGDVYPEVTDNDSVSOVLDMHMLNDEDETSGLRSHLSLVNDTPRLVITGV VFARFMIQSLAAGHAVERGLSRITTVVSGEDSVKLVQQLKVDLHEVRDITHL PFARFELMLIKIIVANAARARNVDIASIFPAKADVADSDHTTELETDLLHKVRLRI LEPYICEVARTDVMHMYVNVQWVIRSRCEIDHITLSKSSRLKIPALCGMCANVD"
BASE COUNT	528 a 348 c 432 g 552 t
ORIGIN	
Query Match	39.8%; Score 665.8; DB 8; Length 1860;
Best Local Similarity	72.6%; Pred. No. 1.3e-179;
Matches 887; Conservative 0; Mismatches 332; Indels 2; Gaps 2;	
QY 288	AAGTGAGCAAGCAACACAATTTCAGTATTGTGTGGACAGCAAGCGG-AATGATTAAATAG 346
Db 240	AAGACTAGGCGCCATACACATTCACATGTTTGTGTGATGAAAGTGAAATGATCAATCG 299
QY 347	GATTGCAAGAGTGTTTGGAGAGAGAGATACAAATATTAGAGTCTTGCTGTGTGCTGAA 406
Db 300	TATTGCGGGAGTCTTGCTGCGAAGAGGGTACATATTAGTCTCTTGCTGTGTGCTTAA 359
QY 407	CAGAGACAAGGCTCTATTCCACATATGTTGTCTGTGGAACTGAAGGGTACTTCAGCAGGT 466
Db 360	CAAGCACAAGGCTCTTTTTCATATGTTGTTTCTGGAAGCTGAAGGGTTCACCAAGT 419
QY 467	CATGAGCAACTCCGGAAGGCTCGTTAAATGTTCTAAGGTGGAAGATATCTCAAGGAGCC 526
Db 420	TATGGAACAGTTTACAAAAGCTTGTAATATGTATCAAGGTTGAAGATCTTATCCAAAGAGCC 479
QY 527	GCAAGTGAGCGGTGAGCTGATGCTTGTAAAAGTGAATGCACATCCAGAATCCAGGAGCA 586
Db 480	ACAAGTTGACAGAGATTAATGCTTATTAATAATCAGCGCGGATCCAAAATACCAGCGCAGA 539
QY 587	GATCATGTGGCGTACGTGTGACACATTTAGAGCAAGATTTAGATATAGGGGAACATGCAAT 646
Db 540	GATTATGTGGTGTGGAGCTGTTTATAGGCGCAAAAATTTGGATATATCTGATCAATCTCT 599
QY 647	GACTATCGAAGTAATCGAGATCTCGAAGAAATGATTCCTGTAGAAAGAAATTTGAAAAA 706
Db 600	GACTATTTAGGTAATCTGGAGATCCAGGGAGAATGSGTGGCTGTTCAGAGGAACATTAAAGTAA 659
QY 707	GTTTCAGATCAGAGAGATTGTAGAGCACAGGAAGATTAGCACTGACAGAGGAAAGATGGG 766
Db 660	ATTTGGAAATTTAGAGAAATTTGCTGTAAGGGGAAGATTTGCTTGAAGAGGAAAAATGGG 719
QY 767	TGCACATCTCCATTTTGGCGATTTTTCAGACACATCTATCCAGATCTTCAAGAGCAAGC 826
Db 720	GGAATCTGCTCTTTTGTGGCGTTTTCAGACACATCATACCAAGATCTTGAAGGTGCAT 779
QY 827	GCTGTGTAGTGTTCGTGAAGTAGCAAAAAAGAGCCATTGTCTCTCAAAAGGAAACATC 886
Db 780	GTCCTCTGTACTATTTCGAGGACAAATCAAGAGGAGCCCTTAATGAGATATCATGTCTAT 839
QY 887	AGCAGGGGAGAGTGTATTCOCGTGAGCCATTTTTCACCCCAAGGTAACATCGATATCT 946
Db 840	GCTGAGGAGATGTCTATTCCTGTGAGCAGATATACAACCTTGGAGTGAAGTCAAGTCT 899
QY 947	CGACCTCACTGGGACTTCTCAGCAGCAAGATACGAGTGATCAAGTGCATACACT 1006

VERSION	ARI94787.1	GI:20241379
KEYWORDS		
SOURCE	Unknown.	
ORGANISM	Unknown.	
REFERENCE	Unclassified.	
AUTHORS	1 (bases 1 to 4895)	
TITLE	Kakefuda,G., Costello,C., Sun,M. and Hu,W. DNA sequences encoding the arabinidopsis acetylhydroxy-acid synthase small subunit and methods of use Patent: US 6348643-A 3 19-FEB-2002;	
JOURNAL	Location/Qualifiers	
FEATURES	1..4895	
SOURCE	/organism="unknown"	
BASE COUNT	1236 a 884 c 1011 g 1703 t	1 others
ORIGIN		
Query Match	25.0% Score 419; DB 6; Length 4895;	
Best Local Similarity	86.7%; Pred. No. 1e-108;	
Matches	507; Conservative 0; Mismatches 0; Indels 78; Gaps 1;	
OY	1	GTCCTTCTCAGTAGCAAAAACCTTCGGCTTCGTCGTCATGTGCGGCATTTCTGTAA 60
Db	717	GCTCTTCCAGTAGCAAAAAAACCTTCGGCTTCGTCGTCATGTGCGGCATTTCTGTAA 776
OY	61	GTTCTTCACCATCTATTTCGCTGCTTGAGATGGCATGTTCGATTTCTTCTGCTCTTG 120
Db	777	GTTCTTCCACCATCTATTTCGCTGCTTGAGATGGCATGTTCGATTTCTTCTGCTCTTG 836
OY	121	TATCCTGACGCGTATTCGTCGCCGGGGAAGATTTCATCTCCGGTATATCTTGGC 180
Db	837	TATCCTGACGCGTATTCGTCGCCGGGGAAGATTTCATCTCCGGTATATCTTGGC 896
OY	181	ACCCTGGCGATGAATGGGTAAAGAATGAAGAATTCGTTGAAGCGTCATGGGAAGA 240
Db	897	ACCGTGGCGATGAATGGGTAAAGAATGAAGAATTCGTTGAAGCGTCATGGGAAGA 956
OY	241	TCTCTGATGCGTCTTTCTCCGAAGCTTCATCTGCACCTCCAAATCGAA----- 289
Db	957	TCTCTGATGCGCTTTCTCCGAAGCTTCATCTGCACCTCCAAATCGAAGCACTGTGAA 1016
OY	290	----- 289
Db	1017	TAAATTTGCTTAAAGTCGTTTTCTTTGGCCTTTGCTTTGATGATTTCTTTGTCAATA 1076
OY	290	-----GTTGAGAGACAACAATTTGAGTATTTGTTGGAGACGAAGCGGATGATTA 342
Db	1077	AAATGAGGCTGAGAGACAACAATTTGAGTATTTGTTGGAGACGAAGCGGATGATTA 1136
OY	343	ATGAGATTGCAGAGATGTTTGCAGAGAGAGATACAAATATTGAGAGCTTGGCTGTGGTC 402
Db	1137	ATAGATTGTCAGAGATGTTTGCAGAGAGAGATACAAATATTGAGAGCTTGGCTGTGGTC 1196
OY	403	TGAACAGAGACAAGGCTCTATTTCACCATAGTTGTCTGTGGAATCAAAAGGTTACTTACG 462
Db	1197	TGAACAGAGACAAGGCTCTATTTCACCATAGTTGTGTGTGGAATCAAAAGGTTACTTACG 1256
OY	463	AGGTCATGAGACAATCCAGAGCTGTTAATGTTCTTAAAGSTTG 507
Db	1257	AGGTCATGAGACAATCCAGAGCTGTTAATGTTCTTAAAGSTTG 1301
RESULT 7		
AF233069/c	6982 bp DNA linear	PLN 04-JUL--2000
LOCUS		
DEFINITION	Albidieria sulphuraria maturase (matk) gene, partial cds: 508	
	ribosomal protein, ribulose-1,5-bisphosphate carboxylase/oxygenase	
	large subunit (rbcL), ribulose-1,5-bisphosphate	
	carboxylase/oxygenase small subunit (rbcs), acetylhydroxy-acid	
	synthase small subunit (ahas), and translation initiation factor	
	IF-3 (lzf) genes, complete cds; and unknown genes; chloroplast	
genes for chloroplast products.		
AF233069		
AF233069.1	GI:8925945	

KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
FEATURES	
source	Location/Qualifiers 1..6982
gene	/organism="Galdieria sulphuraria" /organelle="plastid:chloroplast" /strain="JTEX 2393" /db_xref="taxon:130081" complement(<1..407) /gene="matk" complement(<1..407) /gene="matk" /codon_start=1 /product="maturase" /protein_id="AAE81678.1" /db_xref="GI:8925946" /translation="MKHKDSIFPKKFKLFFKKLLVWIHFSLGARSPYIEMNSKYNIA RADILNYLVDRKYSVDYALLKEDYINKILINRONNAEKRYVAIDLETYDKLITAEKX REAYDNMERRYRKYVLFLDRLRKGHITTKFGGFGL" complement(522..860) /note="ORF339; hypothetical 12.5 kDa protein"
CDS	/note="start=1" /product="unknown" /codon_start=1 /protein_id="AAE81679.1" /db_xref="GI:8925947" /translation="MNTYNINITSKALDLNLSISKLNSEKLYIRIGIKQGGCSGSYST FMTEKSNISNDNDLVYDYDFNLVCDNKSIYLVLGISIDYSSSLDGGRKLIPNAK QTGCCKSKFS" complement(877..1095) /note="similar to L28"
gene	/codon_start=1 /product="50S ribosomal protein" /protein_id="AAE81680.1" /db_xref="GI:8925948" /translation="MDIKMTRCOLLEFTSTNTANKVSHSNHKSRRLOFVNLIQQKKWVS SLKKWTIKRISTSRKMILKNKIDVFL" 1332..2813 /gene="rbcL" /gene="rbcL" 1332..2813 /codon_start=1 /product="ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit" /protein_id="AAE81681.1" /db_xref="GI:8925949" /translation="MSQSLEEKSVOERTRIKNSRYESGVLPYAKMGYNPDYOVKPTD VLAFETPOGPDIPEAAAAGVESTATVTVMPTDLTAADLYRAKAVKVDOVPNN PQGYFAIAYELDFEGSGILANTLASTIGVCPFKAKALRLDMPLPAFYITPCPG ATGVILERLRDLKFRGPLLCCTTKPRKIGLSGKNGRYVYALKGGDLFDVYDDENINSQ PEMKWERVLFVMEAVNKAATAAGEVKGHLNVNTAAMEBTVARAOLAKLGSVTIT DLVIGYAQTAMKWARDNDMLDLHLRAGNSTYSROKNHGMRVCICKMARAGVGVIH HACTVGTGLEGDPIITRGFYKTLLPLERLNDEGLFEFDMNASLRKCVMPVASGQTHA GMHQLHIHYGEDVLOFGGTTIGHDPGIGSDGATVANVALAEMIILARNERNRDLFTLGP ELLREAKNGCALRTALDLMKDITFNTSTDTSDVFETPIANI" 2859..3278 /gene="rbcS" /gene="rbcS" 2862..3278 /codon_start=1 /trans_except=(pos:2862..2864,aa:Met) /product="ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit" /protein_id="AAE81682.1" /db_xref="GI:8925950"
CDS	

[illegible]

Db	4985	TTTGCTAGACGTGATTTAAATATAGATTAGCTTAGCATTGGTCTGCTGCAAAAACGAGCT	1178
Qy	1119	ATTTCACGCATTACACACAGTTATATACCTGCACACAGATGTAATGCGTCAGCAATTTGGTCAG	1178
Db	4925	ATATCAAGAATTACAAATGGTTATCAGGGGCTATATAGAAATTTATAGAACAAATTTCANAA	4866
Qy	1179	CACCTTTCACAAACTCGTATGATGTGCATAGGTCATGATCTTACTTACTATTTGCCATTTTCT	1238
Db	4865	CATCTTTTAAATTAGTAAATGTATTTAAAAATTTACAGTATTAACAAATATTCATGTGTA	4806
Qy	1239	GAAGAGAACGATGCGTATTAAGTTATGCCGTGGAACGCTGCTGCTAGAGAAAGATGTCTCTG	1298
Db	4805	GAAGAGAAATTAACTGATATTAAGTGAAGTGAAGAACTGGAGAGAAATTAAGAAATGATATG	4746
Qy	1299	GACATTCCTACTATTTCAGGGCTTAAACCTGCTGTCAGTATGATCACAACAATTTACTTTG	1358
Db	4745	GAATTCGCCAATATTTTATAGACGACGTGTGTGATTTATCAGAAATCTTTATTTATTTTA	4686
Qy	1359	CAGCTTACTGAGGATCTAGACAGATGTTGTCACCTGCAAGAGTTATTTGAGACCCATATGT	1418
Db	4685	GAGATTAACAGGAGATCCCTGCGCAAAATCCCTCTATTGAAAAATTAATTTGAAAAATATG	4626
Qy	1419	ATATGTGAGTTTGCAGAACCGGTCGTGCTGTCATTTGGCTCGTGAATC	1465
Db	4625	ATTGTGAATTTGCTAGAACAGCAAGATTCCTTTACAGAGAAATC	4579
RESULT 8	AF022186/c		
LOCUS	AF022186	164921 bp	DNA
DEFINITION	Cyanidium caldarium strain R1 chloroplast, complete genome.		
ACCESSION	AF022186	Z36235	Z70297
VERSION	AF022186.2	GI:6465296	
KEYWORDS			
SOURCE			
ORGANISM	Cyanidium caldarium.		
	Chloroplast Cyanidium caldarium		
	Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales;		
	Porphyridiaceae; Cyanidium.		
REFERENCE	1 (bases 130696 to 132364)		
AUTHORS	Vogel,H., Fischer,S. and Valentin,K		
TITLE	A model for the evolution of the plastid sec apparatus inferred from secY gene phylogeny		
JOURNAL	Plant Mol. Biol. 32 (4), 685-692 (1996)		
MEDLINE	97134960		
PUBMED	8980520		
REFERENCE	2 (bases 1 to 164921)		
AUTHORS	Gloeckner,G., Rosenthal,A. and Valentin,K.		
TITLE	The structure and gene repertoire of an ancient red algal plastid genome		
JOURNAL	J. Mol. Evol. 51 (4), 382-390 (2000)		
MEDLINE	20496959		
PUBMED	11040290		
REFERENCE	3 (bases 46857 to 47851)		
AUTHORS	Valentin,K.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-MAR-1996) Institute for Plant Physiology, Justus Liebig University, Heinrich Buff Ring 58-62, Giessen 35392, Germany		
REFERENCE	4 (bases 28701 to 75560)		
AUTHORS	Gloeckner,G., Rosenthal,A. and Valentin,K.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-SEP-1997) Department of Genome Analysis, IMB Jena, Beutenbergstr.11, Jena 07745, Germany		
REFERENCE	5 (bases 1 to 164921)		
AUTHORS	Gloeckner,G., Rosenthal,A. and Valentin,K.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-NOV-1999) Genome Analysis, Institute for Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany		
REFERENCE	6 (bases 130696 to 132364)		
AUTHORS	Vogel,H., Fischer,S. and Valentin,K.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-NOV-1999) Institute for Plant Physiology, Justus Liebig University, Heinrich Buff Ring 58-62, Giessen 35392, Germany		

COMMENT On or before Nov 23, 1999 this sequence version replaced gi:529651,
gi:1240002, gi:2465730.
FEATURES
source Location/Qualifiers
1.164921
/organism="Cyanidium caldarium"
/organella="plastid:chloroplast"
/strain="RK1"
/db_xref="taxon:2771"
join(164808..164921,1..78,462..644)
repeat_region
/note="direct repeat;
repeat unit separated by stem loop"
339..477
stem_loop
/note="stem-loop separates direct repeat unit"
580..780
/note="similar to yeast mitochondrial origin of
replication region in Genbank Accession Number U6902"
gene
986..2041
/gene="psbd"
986..2041
/gene="psbd"
986..2041
/gene="psbd"
/product="unknown"
986..2041
/gene="psbd"
/note="Photosystem II D2 protein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3026.1"
/db_xref="GI:6466445"
/translation="MTAIGREDEGEMFDLLDMKLRDFVIGMSGILLPCAYIAL
GAMFTGTFVSSWYTHGLASSYLEGCNPLTAAVSSPANSMSGSLFLMPGEAOGDFT
WCOIGLMTFTALHSGFLIGFLCELOFEIARLVGLPYNAIAFSGPIAVEVSFLIYP
LGASWEPASPGEVAAIFRPLLEFLOEGHNMTNPHEMGVAGILGALLCAIHGATVE
MTLPERGEASDTPRAFTPROSEETYSMTANTPMSQIIEGVAPANKRMLFLFLLPVYT
GLWVSSIGVGLNLNRAIYDFVSGEIRAEDEPFETFTKNILLNEGIRAMMAAOQOP
HENFVPEEVLPPGNAL"
2031..3413
/gene="psbc"
2031..3413
/gene="psbc"
2031..3413
/gene="psbc"
/product="unknown"
2031..3413
/gene="psbc"
/note="Photosystem II 44 Kd apoprotein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3025.1"
/db_xref="GI:6466444"
/translation="MLFNKDKNNIGRSIESNGFAMWMSGNARLINSGKLGAHVAA
GLIVFWTGMATLFEETSHPIPEKPIYEGMILLPHLATLGNVAPGGEIVNTYPTATG
VTHLVSAVLGRGGITVHSTVGPVLEDSFSFGYDMDKNNKTTIIGHLLIGTAP
LVTKALFTGGITDITWAPGGDIRFTNPTLNALITFYLKSPGEGMIVGNWME
DVIGIHWIGVTCVIGIWHIILTRPFSMARAFVWSGEAYLSYSLGALIMQTALEY
AWNNNTVYSEFGPTPAASQOAFPTFLRQDRLGANIASITQGTGLKLYMRSTFG
EVLIGETMRKFNDRAPMLPELSSNGLDINKIDIOPOERRAEVYMTAPLGSIN
SVGGVATEINSVYVSPRSWMLTTSHTFLPFIFIGILMHAGRARAAAFPEKGINEN
EPVLSMRPLD"
3546..3821
/gene="ycf19"
/note="similar to ycf19 in Porphyra purpurea"
3546..3821
/gene="ycf19"
/product="unknown"
3546..3821
/gene="ycf19"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3024.1"
/db_xref="GI:6466443"
/translation="MKDGDLLIETITIAFLQIYIVILLRMSLGWFPNINWYSQPPYS
LSQSDPYINLPHGVFPSPFLGIDFSPILIGITLIDFTIELLSROLKPL"

gene
3873..4133
/gene="ccrf9"
3873..4133
/gene="ccrf9"
/product="unknown"
3873..4133
/gene="ccrf9"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3023.1"
/db_xref="GI:6466442"
/translation="MNCLPKNNKVFRLNFEKCIDPMCTDNLFEKFPPLYLKNKNGKFT
NKFYKTKNKRYIDILPIFYIYHYILRIGKNQTKDLQITKC"
4189..4272
/gene="ccrf10"
4189..4272
/gene="ccrf10"
/product="unknown"
4189..4272
/gene="ccrf10"
/product="unknown"
4189..4272
/gene="ccrf10"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3022.1"
/db_xref="GI:6466441"
/translation="MTKLKSLNDKKTLISCLNKTIIDIN"
4334..4406
/product="tRNA-Thr"
4450..4971
/gene="ycf3"
/note="similar to ycf3 in Porphyra purpurea"
4450..4971
/gene="ycf3"
/product="unknown"
4450..4971
/gene="ycf3"
/product="unknown"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3021.1"
/db_xref="GI:6466440"
/translation="MERSQKNDNFIDKTFPSIADLIVKILPTNKSKEAFYVYKDMA
AOSGEYVADLACYOALKIEKDPMDKSFILYNIIGHIOASNGHARALEYHESLKF
PNVLOALNNIAVIYHYHGNKLPFGSKLOBAKIMPFKASVYMKRAIKLAVYNTAEQNW
IKITGRITPDIWL"
5004..6476
/gene="atpb"
5004..6476
/gene="atpb"
5004..6476
/gene="atpb"
/product="unknown"
5004..6476
/gene="atpb"
/note="ATP synthase CFI beta chain"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAFI3020.1"
/db_xref="GI:6466439"
/translation="MKPCQLOKOKQTLATAVNNKGVIIQIGPVLDIVFPGDGLPKV
FNAIKINNSNNMVICENVQOLLGDNKVRVAVASTEGLRGASAIIDTGPISIPGKE
TLGRIFENVGEPIDEGPVISNDKLPIDHRPAKFTQLETKPSIFETGIVVDLAPYR
RGKIGIFGAGVGVKTVLIMELINNVAKAHGGSVSGVGCERTREGNDLYOBKRESGV
INEKDNLKSVALCYGQNNPPGARRRVGLTALTAETFRDYNKONVLLFDINIFRV
QAGSEVSALLGRMPSAVGYOPTLGTGEMGALORITSLDGSITTSIQAVYVPADDLTDP
APATTFADATATVLSRALAAAGIYPADVDLSTIMLOPGIVSDEHNTTARKVETL
QRYKEMLQDIIAILGDELSEEDRLIYSRAKIEKFLSOPFVAEYFTGISGKYVSLSD
SIKGFMMILSGEVDNIPQCAFYLIVGRIEBAIDKAKOVERS"
6486..6887
/gene="atpe"
6486..6887
/gene="atpe"
/product="unknown"

[illegible]

Query Match	Best Local Similarity	Score	DB	Length
Matches 276; Conservative 0; Mismatches 182; Indels 6; Gaps 1;	9.0%; 59.5%;	150.8;	DB 1;	300450;
QY 291	GTGAGGAGCACAATTTGATTTGGAGAGCAAGACGAGATGATTATAGATTT	350		
Db 297278	GTTATGAAACATACCTTATGCGTTTGGTGAGATGAGCGGGTGTCCTCAGTCGATTT	297219		
QY 351	CGAGAGGCTTTGCAAGGAGGATACAAATTTGAGATCTTGGCTGTTGGTCT-----G	404		
Db 297218	GCCGGTCTATTATGCGCGCGTGGCTTTTAACTCGAGACCTTGCGGTAGGACTTGGCGAG	297159		
QY 405	AACAGAGACAAGGCTCTATTACCTAGTGTGCTGTGAGACTGAAAGGCTACTTCAGCAG	464		
Db 297158	CAAAATTTGGCATCTCCCGATTATACATATGGTGGTGGCTGAGATGATGCTGTCAATTGAGCAG	297099		
QY 465	GTCATCGAGCAACTCCAGAAAGCTCGTTAATGTTCTAAAGTTGAAGATATCTCAAGTGAG	524		
Db 297098	TTGACCAGCAACTCTACAGCTGATCAATGATGCTCAAGAGGTGACAGATATTACCACAATTT	297039		
QY 525	CCGCAAGTGGAGCGTGAGCTGATGCTTTGAAAGTGATGACATCCAGAAATCCAGGSCA	584		
Db 297038	CCCTGTGTAGAAACGGGAGCTGATGCTCTGTGAAGGTGAATGGCACCTGCCAGAACCCGCTCT	296979		
QY 585	GAGATCATGTGGCTCTAGTTGACATCTTCAGAGCAAGACTTGTAGATATGACCGAACATGCA	644		
Db 296978	GAAATCTTGGAATTTGGTCAAAATTTTCCGCGCTAAGGTCGTGGATGTCTCCGAGCATTCG	296919		
QY 645	TTGACTATCGAGGTAACTGAGAGATCCTGGAAAATGATTTGCTGTAGAAAAGAAATTTTGAAA	704		
Db 296918	CTCATCATCGAGGTCCTCCGGTGACCCCTGGCAAAATGTTGGCCCTAGTGCAAATGATGTGAAT	296859		
QY 705	AAGTTTCAGATCAGAGAGATTGTAAGACAGGAAAGATGACACT	748		
Db 296858	AAATTTGGGATTCGCGAAATTTGCCCGCACCGTAAATTTGGCCT	296815		

RESULT 10
 AE013291/c
 LOCUS
 AE013291
 11318 bp
 DNA
 linear
 BCT 17-MAY-2002

DEFINITION	Methanosarcina mazei strain Goel, section 73 of 379 of the complete genome.		
ACCESSION	AE013291 AE008384		
VERSION	AE013291.1 GI:20905057		
KEYWORDS			
SOURCE	Methanosarcina mazei Goel.		
ORGANISM	Methanosarcina mazei Goel.		
REFERENCE	Methanosarcinales; Archaea; Euryarchaeota; Methanococci; Methanosarcinales; Methanosarcinaceae; Methanosarcina.		
AUTHORS	1 (bases 1 to 11318) Depemneier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A., Brueggemann,H., Lienard,T., Christmann,A., Boemcke,M., Steckel,S., Bhattacharyya,A., Lykidis,A., Overbeek,R., Klenk,H.-P., Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.		
TITLE	The genome of Methanosarcina mazei: Evidence for lateral gene transfer between Bacteria and Archaea		
JOURNAL	J. Mol. Microbiol. Biotechnol. (2002) In press		
REFERENCE	2 (bases 1 to 11318) Depemneier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A., Brueggemann,H., Lienard,T., Christmann,A., Boemcke,M., Steckel,S., Bhattacharyya,A., Lykidis,A., Overbeek,R., Klenk,H.-P., Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.		
AUTHORS	Direct Submission Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Grisebachstrasse 8, Goettingen 37077, Germany		
TITLE	Location/Qualifiers		
JOURNAL	1. .11318		
FEATURES	/organism="Methanosarcina mazei Goel"		
SOURCE	/strain="Goel"		
gene	/db_xref="taxon:192952"		
CDS	931..1776 /gene="MM0663"		
gene	931..1776 /gene="MM0663"		
CDS	/codon_start=1 /transl_table=11 /product="putative dihydroorotate dehydrogenase electron transfer subunit"		
gene	/protein_id="AAM30359.1"		
CDS	/db_xref="GI:20905058"		
gene	/translation="MVYRILEKEEIIAPSVHRMVIDAPYAKAKAGOFILRIDERGE RVPILTADEKGTGTVIYFOEMKTKTQOLAKLSAGDSLEDVPGITPADIRKIGTV ILVGGVGAVPYPOAKAVYVNGKVISITGARNKDILILEDPMKASSELYIATDGG SKGHRGVTIDMKRILDSGEKARVYIIGPITMKRVGAVSPYVEILVLSNSTIADV GTCGCGCRVTVYGGTFTCYDGPEDARKVDVQIMNLAMYRGEETEATEKYEEKC RCGHL"		
gene	1795..3204 /gene="MM0664"		
CDS	1795..3204 /gene="MM0664"		
gene	/EC_number="1.4.1.13"		
CDS	/codon_start=1 /transl_table=11 /product="glutamate synthase [NADPH]"		
gene	/protein_id="AAM30360.1"		
CDS	/db_xref="GI:20905059"		
gene	/translation="MOELHGEKTKGAKKERPPMPPOAGERRKNFEVALGYTKEDAL AEASRLACKEPKCVGEPVNDIPGFIOLICEENFEAGIERIKATNALPAICGRPAD OESOEALHCVLGKGQPAVIGLERFCADYERKGVSPQVMPGKRAVAVGSGPAG LTAADLAKIGHKVTIIFESIKRAGSVLSGIPERIPKEIVROELIYEKIGVEVRN YIIGIKITLDELCDPDAVFLCTGAGLSFMSICEENLNGVYSANEFLTRVLMKAYD PEYDITVRKGNVAVGGGVNAMDARSALRIGADEVSLIVRIGDEDEMARRELEIHA KEEGILTFRLTNPRILGDEKFNVTGVCIRIMELGEPKSGRRSPVPVGSSEFLPAD VVVAIGTSPNPIIFKSGSGLEONKRGTVVADEETGATSKCGVAGGDVVTGAATVIS AMGAKKAALDEVILKEK"		
gene	complement(3320..4189)		
CDS	/gene="MM0665"		
gene	complement(3320..4189)		
CDS	/gene="MM0665"		
gene	/codon_start=1		
gene	/transl_table=11		
CDS	/product="hypothetical protein"		
gene	/protein_id="AAM30361.1"		
CDS	/db_xref="GI:20905060"		
gene	/translation="MEKNIFANLGSKYSADADEINTSASLINSAEITRIMSVDQVSS TYSNAKINLMASETDPEALKILNNASSDYDEEIDITFKTLAEGGLDRVHLSLI TAMEHSDKSMAYAAVSKERNLSRKELININELNESAASISAKERVYIDPESVPIEQ KSSILINLNDLEASETMSELKROMSGMPYWDGYVCLNSGLEYGDAEEMGKADEFG KASDKSESOKLETILKDSXSESVTALIEICGILTQAKDPLHTEACGRVNEKRYI QANAEFNVSYYY"		
trna	complement(4987..5058)		
gene	/product="trna-Arg"		
CDS	5520..5960 /gene="MM0666"		
gene	5520..5960 /gene="MM0666"		
CDS	/codon_start=1 /transl_table=11 /product="putative nickel-responsive regulator Nkr"		
gene	/protein_id="AAM30362.1"		
CDS	/db_xref="GI:20905061"		
gene	/translation="METELMRIGVSLPDTLLCKEPDIEIKRQYSSRSEAIKDSISYI STNEMGNKIGHVGTIVIIDHTQGLSNVLTIDQHHSHLIRSHIHYPQREDCFE IVLIEGKEIITELAEALMAIKVLSKILITESNETDEPSIV"		
gene	complement(6758..7300)		
CDS	/gene="MM0667"		
gene	complement(6758..7300)		
CDS	/gene="MM0667"		
gene	/codon_start=1		
CDS	/transl_table=11		
gene	/product="Nitroreductase family protein"		
CDS	/protein_id="AAM30363.1"		
gene	/db_xref="GI:20905062"		
gene	/translation="MPASGENKVIDIILNRSVREFPNDRPISKEIKNTILNAGRAP SGLNQPWRIVIRDEPTILIKISECHYSIIIEGAPLLAFLDPTNSYNRKQDVAIG ASTQMLLACCDLGGAVMLGEILKONKRVNSILHCPPELDIMAVLAIGEPAGKRTS SRPLSEILFVNEKYGCKWE"		
gene	complement(7560..8567)		
CDS	/gene="MM0668"		
gene	complement(7560..8567)		
CDS	/gene="MM0668"		
gene	/EC_number="1.1.1.86"		
CDS	/codon_start=1		
gene	/transl_table=11		
gene	/product="ketol-acid reductoisomerase"		
CDS	/protein_id="AAM30364.1"		
gene	/db_xref="GI:20905063"		
gene	/translation="MAIIIVDEDTTFDALDKRTIAVMGYSGGHAHARNLHESGLNVI VGLROGSSMAKAESDGLKVMTVDDASRAADYIMLLPDERQAAVYSOIEENLKAGD ALVFAHGNHYNQIVPPKDVDPVAVPAGPHIYRTYTGIGVGLIAYVQDATGK ARDLALSYAKGIGATRAGYETTFTRPEETPTDIFGQVYDLCGLSLIKTAPEVLYEAG IOPENAVFETCHEVKLIVDLITYEGGLERMHVSNTAEYGGTVPVRYINEYSREAMY EALERIIONGEFAKEFVLDEGMVNHVPLKAMEROEKBOLEVGKQIRANIPWLNRKIID D"		
gene	complement(8961..9446)		
CDS	/gene="MM0669"		
gene	complement(8961..9446)		
CDS	/gene="MM0669"		
gene	/EC_number="4.1.3.18"		
CDS	/codon_start=1		
gene	/transl_table=11		
gene	/product="Acetolactate synthase small subunit"		
CDS	/protein_id="AAM30365.1"		
gene	/db_xref="GI:20905064"		
gene	/translation="MKHTLVAIVENKSGVLSRFVASFSGRYNIDSLAVGTEDPDIS RTIIVHGDDHLEQVTKQNLKILIVIKVSDIGSDDAVERELALIKVSADVPTRAEII QIANIFRAIVVAPKSMIVETVGDGKIQALEKLLRQGIKEMVRTGIIAIVRQPK V"		
gene	complement(9443..11137)		
CDS	/gene="MM0670"		
gene	complement(9443..11137)		
CDS	/gene="MM0670"		
gene	/EC_number="4.1.3.18"		

		/codon_start=1	
		/transl_table=1	
		/product="Acetylacolate synthase large subunit"	
		/protein_id="AA030366.1"	
		/db_xref="GI:20905065"	
		/translation="MTTGSPTDKVNGAKALIKLEKEGVDTLFGYPGGIIIPFYNELYDR SLLHILVRHEQAAAHAADGYARATGTGCVSTNSGATNLTVMGLATAMDSVPYVA LTGVSPTSLLIGNDAFOEADITIGTMPITKNNYLVDQDEIPRIYKEAFHLASTRPPGR VLIDPVDVNIETIDLTPPERVELRGIRPYTKNTQOIKRAEELIANSCTPIYAGGG VISSNAAEIVIEFAETIKAPYTEVTLTMGISIPTEHPLYVGLGHGCKRYANVAIQESD LIILVGAFFDFDRVTGLKESEFNPARNRVHIIDVDPLEISKNVAVHIPVGDAKOYLSLD RYVCCSAEWIEKVYNLMKREPLNVRECRTPIMPOVEIESVCDAIIVTEVGQH OMAAOFEPKYSKPRTFLISGLGTGMYGFPAMAKAKGRDPKTINTINLAGOSSPOMNSO ELAVIWNNDIPYVSVIILNGLGNVRMOELEFYTRNRVSYPIFIKSVDPLYKAELYGAL GLRERSEVERPAILEEVNRSRPVLEYICECAENVPMYPAGAINEIIDLEEHR"	
BASE COUNT	3290 a	2435 c	2450 g 3143 t
ORIGIN			
Query Match	9.0%	Score 149.8:	DB 1: Length 11318:
Best Local Similarity	59.4%:	Pred. No. 2.2e-31:	
Matches 275:	Conservative 0:	Mismatches 182:	Indels 6: Gaps 1:
Dy	293	GAGGAAGCACACAAATTTTCAGTATTGTTGGTAGACGCAAGCGGAATGATTAAATAGGATTGC	352
Dd	9447	GATGAACAATACACTTGCGAGCTCCTGTGGAAATAAATAGTCGGTGTGCTCCAGAGTTGC	9388
Dy	333	AGGAGTGTTTGCCAAGAGAGATACATATTTGAGAGTCTTGCTGTTGGTCTGAACAGAGA	412
Dd	9387	TTCCTCTTTTCAAAGCGCTGGATACAAATATCGACAGCCTTGTCAGAGGGAGTAACCGAGA	9328
Dy	413	CAAGGCTCAT-----TCACCATAGTTTGTCTGTGGAACTGGAAGGTAATCTCACCAGGT	466
Dd	9327	CCCTACATCTCGAATATPACCATTTGTTGTGCATGGGATGACCATGTGCTGAACAGGT	9268
Dy	467	CATCGAGCAACTCCAGAAGCTGTTAATGTTCTTAAAGGTGAAGATATCTCAAGAGAGCC	526
Dd	9267	TACAAACAGCTCAACAAAGCTATATTGATTATCAAACTCTCGGATATTCGGAAGCGATGA	9208
Dy	527	GCMACTGGAGCGCTGAGCTGATCTGTAAAGTGAATGCACATCCAGATCCAGAGGCGAGA	586
Dd	9207	TGCCCTGTAAGGAAGAACTTGCCCTTATFCAGAGTCTCAGGTGATGTGGACACAGGCGAGA	9148
Dy	587	GATCATGTGGCTAGTTGCACATTTAGAGCAAGAGTTGTAGATATAGCGGAACATGCATT	646
Dd	9147	GATTATCCAGATTTGCAAACTATTTTAGGCGCCGAGTAGTGTGATGTGGCGCGAAGTCCAT	9088
Dy	647	GACTATCGAGTAACTGGAGATCTCGAAAAATGATTTGGTGTGTAAGAAATTTTAAAAA	706
Dd	9087	GACTGTTGAGGTAAACAGGTGATGAAGCGAACAGATCCAGCGAATCGAAAAACTATCTCCGCA	9028
Dy	707	GTTTCAGATCAGACAGATTGTAAAGCACAGAAAGATAGCACTG 749	
Dd	9027	GTTTCGAGATCAAGAAATGGTCCAGACACAGCAAGATTTGCAATG 8985	
RESULT 11			
LOCUS	PU38804/c	191028 bp	DNA circular PLN 27-MAR-1998
DEFINITION	Porphyra purpurea chloroplast, complete genome.		
ACCESSION	U38804		
VERSION	U38804.1	GI:1276652	
KEYWORDS			
SOURCE	Porphyra purpurea.		
ORGANISM	Chloroplast Porphyra purpurea		
	Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;		
	Porphyra.		
REFERENCE	1 (bases 1 to 191028)		
AUTHORS	Reith,M.E. and Muniholland,J.		
TITLE	Complete nucleotide sequence of the Porphyra purpurea chloroplast genome		
JOURNAL	Plant Mol. Biol. Rep. 13 (4), 333-335 (1995)		
REFERENCE	2 (bases 1 to 191028)		
AUTHORS	Reith,M.E.		

FEATURES	source
gene	Direct Submission Submitted (17-OCT-1995) Michael E. Reich, Marine Biology Section, NRC Institute for Marine Biosciences, 1411 Oxford Street, Halifax, Nova Scotia B3H 3Z1, Canada Location/Qualifiers 1. .191028
CDS	/organism="Porphyra purpurea" /organelle="plastid:chloroplast" /strain="Avonport" /db_xref="taxon:2787"
gene	181. .1053 /gene="chl"
CDS	181. .1053 /gene="chl" /codon_start=1 /product="protochlorophyllide reductase iron-sulfur ATP-binding protein" /protein_id="AAC08073.1" /db_xref="GI:1276653"
gene	/translation="MKLAVYVGKGGIGKSTTSCINISVALSKRGKKVLDICGDPKHDTFF TLTGELIPTIITDLOAKDYHEEDVPDVLYIKGYGVDCVAGGPAGACGGYVGE TVKLKLEPLIADEVDIIILFDVLGDVCGFPAALANYADYCLITDNCFDLAFANRYLA ASVEKARTSLRLAGLVGRKTRDRIDIXYICDPMPVLEVLPLEDIVSRVKGT LFEMASERTNLAYVCDDYYLNADQLARPQGVVKEAADRELPSLSDFVLNPKSNTT KSTADEBELDLAMV"
CDS	1095. .2402 /gene="chlN" /gene="chlN" /codon_start=1 /product="protochlorophyllide reductase chlN chain" /protein_id="AAC08074.1" /db_xref="GI:1276654"
gene	/translation="MSIVQSDFALFECEETGNHTFCPIISCVSWLYOKIEDSFELYIST KTGCYFLONAGWVIIFAPRPMALKEEGDISALINDYGELRLGLCIKKDRNSVIE WICGTTEIITKMDELGTAEPKLEAIRPIYVARANGDIYAFTGEDTVLIAMOROS HHYTENTPNHYPLVLRGSLEPDVVYTOULVMELKOGITVSGMLBSKRTELPIVOEGR VAGNEPLSLRTATTLMRRRKTKLLGAPFPICGPDOTRAMIERKICVSFIKEGUALERK DIMQSLVEDYIALIRGKSVFFMGDNILEVSLARLETGCMYEVYEIGIMDKRYOAAEL ALLSTCDMMVMPTIVEKPKDNYNOLDRIIDLCPDLVTGMHANPLEARGINTKWS VELPEAIOHGFTNMRDILLVTRPLRNLSISELGWDIYSKG"
CDS	complement(2454. .3929) /gene="Orf491" complement(2454. .3929) /gene="Orf491" /codon_start=1 /codon_start=1 /protein_id="AAC08075.1" /db_xref="GI:1276655"
gene	/translation="MNFITDRLRLKSRYPILINTREBERLEYINKOLFCDENKV YCMFVDGYTSNPNNDNGYAKNNPLDALEFLETDTPYLNFIILDGFSFNFDLGLIK LRNLSQIITKIOSKNIIIVCSKNVPIPLSDVTTLPDLPLSEIKKEIRTSIALSI VLDELVNNAKSCGQSVIRIKTIKITTAOYQOLDSRSLPIIEEKROLINOTRLL EFYVFKATEDIGDGLAKLSWLKKRSRSPKSNQSYGIPPKGILLVIGTGTSLTA KAINDMTPLRLRIDICKLRGLYGESBRREMYITAEGSPCVLMIIDEIDKAFSMU YSODSOSTSAKVFGETITWLSSEKTPPVVAATANTTONLSEMLRKGRFDIFPLDIP NNERSELFIQHILARIPKSNOWNYNIQSLSLCKFGAELAEQAIVESMHYASERE ESTIDIKIALEOFPLAYTDEQVESIQAWAGNRARNAST"
CDS	complement(4047. .4187) /gene="rpl34" complement(4047. .4187) /gene="rpl34" /codon_start=1 /product="50S ribosomal protein L34" /protein_id="AAC08076.1" /db_xref="GI:1276656"
gene	/translation="MTKGTLOGSKRRKIRISGFARRKKTSPGSRILNERRRKRAKIM AS"
CDS	4333. .4854 /gene="ycf37" 4333. .4854 /gene="ycf37" /note="hypothetical chloroplast ORF 37." /codon_start=1

gene /protein_id="AAC08077.1"
/db_xref="GI:1276657"
/translation="MLVLPPIFYLSITLFLPFLSWLISQCLKTIFLLESQFOYFLDK
SNDNLDGEDSEFAFSKCIARKEFTFRAIIESHVLKKNLSLESPENAAIIAKYINMG
FYTYADOKTKLAKNYERAIIEVDGNVYIALNNLNAKIIYEDPKNLKALVDKVLNAK
SNKIANTHKKRTK"

gene complement(4889. .4962)
/gene="trnp"
complement(4889. .4962)
/gene="trnp"
/product="rRNA-Pro"
complement(5020. .5544)
/gene="ORF174"
complement(5020. .5544)
/gene="ORF174"
/codon_start=1
/protein_id="AAC08078.1"
/db_xref="GI:1276658"
/translation="MIEKTFQNSTVGESEYNNFKFLFLDKTFDKIELKDYLSNT
KNIINLELEQCDSYGVWKRPLKVKYIALKHSIIISLIQKNDSTPLVAFARATSDN
GENAIIIMVYVHPIDFQGLGKVVHMLIKQLPQAEISTITLFAEPDVISFYRKLGRT
KDPDGVKGMFVPR"

gene 5807. .6364
/gene="psaP"
5807. .6364
/gene="psaP"
/codon_start=1
/transl_except=(pos:5807. .5809,aa:Met)
/product="Photosystem I reaction centre subunit III"
/protein_id="AAC08079.1"
/db_xref="GI:1276659"
/translation="MKKATFTICLLAVLVSNPITVNAEVAAGLIPCKDSAAFNKRVN
SAGKIQALFLAKYDANTPPALANKQIEKTKTFATYGRAGLCTDGLPHLISDGRMS
RAGDVPFPGFLVLTGIGWGRGYLLSVAKTSKPKPEKIIIDVPLAVFMSSGFAM
PLAAQFESSQLAPNDITVSPR"

gene 6377. .6505
/gene="psaJ"
6377. .6505
/gene="psaJ"
/codon_start=1
/product="Photosystem I reaction centre subunit IX"
/protein_id="AAC08080.1"
/db_xref="GI:1276660"
/translation="MNNFTKYLSTAPVIGVLMWTFAGFIIELNREFPDVLFYL"
6660. .7145
/gene="apcD"
6660. .7145
/gene="apcD"
/codon_start=1
/product="Allophycocyanin gamma chain"
/protein_id="AAC08081.1"
/db_xref="GI:1276661"
/translation="MSIVSQIINADDELRYPTIGELQSIQNYLITGSSNRIATITIR
DKKPIIKASQIQFLHPEYIAPGNAAGSRKRSCLRDYGWTLRLITGVGLAGDDSD
SETIIGIIGVEYMSNLGVPITIGMLDAIQCLKEASLEMLGQDAKIIISPEVDYIIRGM
S"

gene 7473. .8453
/gene="fabH"
7473. .8453
/gene="fabH"
/codon_start=1
/product="beta-ketoadcyl carrier protein synthase II"
/protein_id="AAC08082.1"
/db_xref="GI:1276662"
/translation="MGVHIISTGSSVPEVSVENQFEDDIETSDHWISTRTGKKRHL
AASISIKLAAEANDALSKASINAEIDILIIATSTPDDLFGSASOLAEGATSS
TAFDITGACSGEIIALVTASQFIQAGSYNKVLVYADTMSRWIMDSRSSCIILFGDA
GAVLITGSSINSILGFKCTGQGLNSHLQNLNPSDSQSGELTTPVPGRDSIIRMNK
EYKRAVQVPIPLKNCINDVINSIDEDWELIQAANRIIEAATLSLISPLSMITIN
LENYGTSNAPSPALDEAIKEKKIQPQVYVVLGFGAGLTGCAIVLKWQ"
complement(8465. .8536)
/product="rRNA-Leu"
8735. .8854

gene 8735. .8854

[illegible]

Db 33093 AATTGGCATTATTGAATAGTACGAGCTGGAAGATTTCGTAGTAG 33045

RESULT 12
D63676/c
LOCUS
DEFINITION
Cyanidium caldarium trnC, trpA, trnT, rps4, trnM, rbcL, rbcS, cfxQ,
trnC, trnL and tlyH genes, complete cds.
ACCESSION
D63676
VERSION
D63676.1 GI:2541887
KEYWORDS
tRNA-Leu; tRNA-Cys; tRNA-Thr; tRNA-Lys; tryptophan synthase alpha
subunit; thioedoxin m; ribosomal protein S4; ribulosebiphosphate
carboxylase small subunit; ribulosebiphosphate carboxylase large
subunit; acetoacetate synthase; cfxQ protein.
SOURCE
Cyanidium caldarium (strain:RK-1) DNA.
ORGANISM
Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales;
Porphyridiaceae; Cyanidium.
REFERENCE
1 (sites)
Ohta,N.
TITLE
Analysis of a plastid gene cluster reveals a close relationship
between Cyanidioschyzon and Cyanidium
J. Plant Res. 110, 235-245 (1997)
JOURNAL
2 (bases 1 to 7550)
Ohta,N.
TITLE
Direct Submission
JOURNAL
Submitted (31-JUL-1995) Nijl Ohta, Waseda university, Research
center for Human Sciences; 2-579-15 Mikajima, Tokorozawa, Saitama
359, Japan (Tel:0429-47-6835, Fax:0429-48-4314)
location/Qualifiers
1. 7550
/organism="Cyanidium caldarium"
/strain="RK-1"
/db_xref="taxon:2771"
12. .83
/gene="trnC"
12. .83
/gene="trnK"
108. .836
/product="tRNA-Lys"
108. .836
/gene="trpA"
108. .836
/gene="trpA"
/codon_start=1
/product="tryptophan synthase alpha subunit"
/protein_id="BAA22823.1"
/db_xref="GI:2541888"
/translation="MFIAYLTAGAPDLNTTKQALNLANGDADVIEIGVPYSDDLADG
MILQKASQALNKRFLQDLMLAEILPVPVILAYNOIFHGVGEKWTLLVNLK
VVALIYVDLPYEAEALILRAACARHNLHMTLISPTPKYRAKOLALACDDMIYLSRT
GYGVADHFDNDIOPNMIIELEKQVTTPIALGFIHQKQQLQVKSAGADGVIIGTACM
OILLEKGVOLTEWISAMKSSSP"
800. .2023
/note="ORF407"
/codon_start=1
/protein_id="BAA22824.1"
/db_xref="GI:2541889"
/translation="MDKDEKILIPYGATLAEISKONTPIPEFIYVFTQGMITLN
DILDELKLFQAKYEFKMKQMSSESDVDIAIMAYERLDEGEVMSQWSTI
STIDRAEVLQMPKLAALVHLTOQIYDKDILHHPGASMAVYESRDKALGLVYTHLLRK
SLIKGDHLEAGVSGKIRALNANVQKQPYTEAKAGIHLTGNOVLGATAIINKGAK
EKENTSNORIHILCHSKGNLIVSESOSGLENCYLVQEKWELLSAEGSIRMODIR
LAODTNSILVAMOTMRKROVVEAYDEFOILITRLOQIEAOPISITISRAVVKOYFR
OGTIAGCTVTKGCFQGMVCOILRGHNLKATAHINIKRIHKEVKEVKASQEGISIMK
EMNDQVDEIVLIN"
EMNDQVDEIVLIN"
complement(2023. .2095)
/gene="trnT"
complement(2023. .2095)
/gene="trnT"
/product="tRNA-Thr"
2096. .2674
/gene="rps4"
2096. .2674
CDS

/gene="rps4"
/codon_start=1
/product="ribosomal protein S4"
/protein_id="BAA22825.1"
/db_xref="GI:2541890"
/translation="MSRYLGPRVRIIRLIGALPALTNKSPKKRTIAPGEHAHKRLS
EFAVLOEKOKLQYVYGITNSOLARYFRQARSKSTGTALITMLETRLDHLVYRAG
APPLLAAROLVNHGHVYKNGKKTITASNAECENAHIEVSSSPKPEPLPYLDLSN
GTLVTOPQOKEMLAFFVNELLVETRYRGA"
complement(2671. .3432)
/note="ORF253"
/codon_start=1
/protein_id="BAA22826.1"
/db_xref="GI:2541891"
/translation="MDDYIFPEYRLSEPOPTWQSIAMQVYSPVOIOMLLSROYOSI
HLPIFKSKSGVLDYOFHOFKSNPKIHDAFROSLEACPITYVNVNPKOELLISPRK
MSEIQAPTFMNMYVDCVFRQSPSTLYLFFIPDDONQTLQSVLNNPFERKNSVOE
GDLFTFLKFNRAADVVLVPHLKQCMQVCPGTVPFRVQADREMLELSKONAOQFV
KMKRLSCQDVLSLEADANQARCVLPIS"
complement(3608. .3913)
/gene="trnM"
complement(3608. .3913)
/gene="trnM"
/codon_start=1
/product="thioedoxin m"
/protein_id="BAA22827.1"
/db_xref="GI:2541892"
/translation="MQIDELSFDTVEVLSHOLVLDENAPMCGPCRMIAPIILDEAOE
LDIKVKKLMTDNPMLATEYGIKRSIPTMLFKGGRDITIGAVPRSTLMTIEYTL"
4046. .5512
/gene="rbcL"
4046. .5512
/gene="rbcL"
/gene="trnL"
/codon_start=1
/product="ribulosebiphosphate carboxylase large subunit"
/protein_id="BAA22828.1"
/db_xref="GI:2541893"
/translation="MAOSVOERTRLNKRRESGVIPYAKMGYMDPNYVYKSDVLALE
RYTPQGVDPTEAANAAGESSATMTVWTDLTACDYYRAKAVKVDAPVADQYF
VYIATDIDLEEGSIALNLASTIGAVFGKATKGLRDEMRJLPSFKTFQCPAGTGI
VERERLNKGRPLDGTCKPKLGLSKNKGRAVYELSKGLDLODEINISQPRRW
RDFLYVMGVNRRASASGEVGSYLVNVAAMEIYKRAEFAKGVITIMIDLVIG
YTAIQSMAYVSRNNMIILHHRAGNSTYSRQNHGINFVIRKMKMAVGVGDIHAGTG
VGLKGDPEVYIIGFYNTLLPKLEINLPGLEFFEDMWSIRKVMVVASGSHIAGOMQ
LIHNGEVDVLOFGGTTGHPDGIQAGATANVALESMILANEGRDYFNEGPQLLRD
AAKNGCPLOTALDLMKDISFNYSNTSDTSDFVETPIANV"
5518. .5934
/gene="trbcs"
5518. .5934
/gene="trbcs"
/codon_start=1
/product="ribulosebiphosphate carboxylase small subunit"
/protein_id="BAA22829.1"
/db_xref="GI:2541894"
/translation="MKVTOGFPCYLPDLTDEQIRKQLOYAINNGMALSYETTDPPHR
NSYEMWGLPLDPDKDPATVMEFQVEQRCROYNYIKVLAENSTGIESTALSTFIVNR
PVHPEGFVLRKIESNDRVQRTYVHSVATEKPEGSRR"
5991. .6851
/gene="cfxQ"
5991. .6851
/gene="cfxQ"
/codon_start=1
/product="cfxQ protein"
/protein_id="BAA22830.1"
/db_xref="GI:2541895"
/translation="MKVNLQTEFFKQTOIQOVLDDLDRELIGLQAVKTRIRIETALLV
DLRLQKLGSSNSPGLHMSFTSGPGTKTVAATKMAADILYRGLYIKGHLITVTPDDL
VGOYIGHTPAKTKOYLKNAIMGVLPIDEAYVYLRPDNEBDYGAEAIEIHLLOYMEMORD
DLVITFAGYKEMKDRFYASNPGLSSRVANHVAMPYTPBELMICKIMLOEOQYOFTP
EAEVFLQYIQRRMQQPHFANAKRSVARNALDRARLKOANRIFPTPQNLTKIDLVIYIA
EDILKCRLEQ"
6862. .6933
/gene="trnC"
/gene="trnC"

tRNA	6862..6933
gene	/gene="trnc"
gene	/product="tRNA-Cys"
tRNA	6938..7023
gene	/gene="trnl"
gene	6938..7023
gene	/gene="trnl"
gene	/product="tRNA-Leu"
gene	complement(7024..7524)
gene	/gene="llvh"
CDS	complement(7024..7524)
gene	/codon_start=1
gene	/product="acetylactate synthase"
gene	/protein_id="BA022831.1"
gene	/db_xref="GI:2541896"
gene	/translation="MKHTLSVLEDEAGVLRINAGLFRRSRNIESLAVGPDEOYGIS
gene	RITMVPDDRIEOLMKOLKFIPIFVENLTOVPCVERELMKVANSDFREIL
gene	DMANIFRARIIVDIAIEDLMLEVTGDPKRMVLEOVLAKEGLVEVARTRGKIALKRSSKY
gene	NTEWLR"
BASE COUNT	2486 a 1367 c 1524 g 2173 t
ORIGIN	
Query Match	8.8%; Score 147.2; DB 8; Length 7550;
Best Local Similarity	56.0%; Pred. No. 1.2e-30;
Matches 278; Conservative	0; Mismatches 218; Indels 0; Gaps 0;
999	CTACTCTATCATTTGGCTTAAATGATATTCAGAGGCTTCTTAATTTGCTGACTGTT 1058
7518	CATCTATTATTCGTATTAGTAGAGTGAAGCGAGGTCTTAAGTGAAGGGGTTTA 7459
1059	TTTCGCTGAGAGGGATACATATCCAGAGCTTGCCGTAGACATGCTGAACCAAGGC 1118
7458	TTTGGCCGACGATAGCTTTAACTAGAACTTTAGCGGTAGACCCCGCAGACAGTGGC 7399
1119	ATTTCACGATTACAACTATTAACCTGCAACAGATGAATCGCTGACAAATTTGTCAG 1178
7398	ATCTCTAGAAATTAACATGCTGCTGCTGAGATGACAGAACATGACAAATTTGATGAAA 7339
1179	CAACTTTACAATCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1238
7338	CAATTTATACAAATTTATCCCAATTTTCAAGTAGAAATTTGACACAAAGTCCCTGTTA 7279
1239	GAAAGAGAACTGATGCTGATTAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1298
7278	GAAAGAGATTAATGTTATGAAGTGAAGCTACTGACACACAAAGAGAAATCTTA 7219
1299	GACATTGCTAGTATTTTCAAGGCTAAAGCTGTTGACGATCTGATGACACAAATTTACTTTG 1358
7218	GATATGCGCCAAACATCTTTCGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 7159
1359	CAGCTTACTGGGATCTGACAAAGATGTTGACACTGCAAGGTTATTTGAGGCCCTATGCT 1418
7158	GAACTGACTGGGATGATCCCGCAAAATGTTGCTTGAACAAGTATTTACTAAATTTGGA 7099
1419	AAATGAGAGTGGCAAGAACCGGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1478
7098	TTAGTGAGAGTGGCTAGAACCGGTTAAATTTGCTTTAAACGATCATCAAAAGTGAATCT 7039
1479	AAGTACCTTCTGAT 1494
7038	GAATGCTTACGTTAAT 7023
RESULT 13	
AE011091/c	11759 bp DNA linear BCT 03-Apr-2002
LOCUS	AE011091
DEFINITION	Methanosarcina acetivorans str. CZA, section 436 of the
ACCESSION	complete genome.
VERSION	AE011091 AE010299
KEYWORDS	AE011091.1 GI:19917861
SOURCE	Methanosarcina acetivorans CZA.

ORGANISM	Methanosarcina acetivorans CZA
REFERENCE	Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
AUTHORS	Methanosarcinaceae; Methanosarcina.
1 (bases 1 to 11759)	
Galagan,J.E., Nusbaum,C., Roy,A., Endrizzi,M.G., MacDonald,P.,	
FitzHugh,W., Calvo,S., Engels,R., Smirnov,S., Atwood,D., Brown,A.,	
Allen,N., Naylor,J., Stange-Thomann,N., Dearellano,K., Johnson,R.,	
Linton,L., McEwan,P., McKernan,K., Talamas,J., Tirrell,A., Ye,W.,	
Zimmer,A., Barber,R.D., Cann,I., Graham,D.E., Grahame,D.A.,	
Guss,A., Hedderich,R., Ingram-Smith,C., Kuetner,C.H.,	
Krzycki,J.A., Leigh,J.A., Li,W., Liu,J., Mukhopadhyay,B.,	
Reeve,J.N., Smith,K., Springer,T.A., Umeyan,L.A., White,O.,	
White,R.H., de Macario,E.C., Perry,J.G., Jarrell,K.F., Jing,H.,	
Macario,A.J.L., Paulsen,I., Pitchett,M., Sowers,K.R.,	
Smanson,R.V., Zinder,S.H., Lander,E., Metcalf,W.W. and Birren,B.	
The Genome of M. acetivorans Reveals Extensive Metabolic and	
Physiological Diversity	
Genome Res. 12 (4), 532-542 (2002)	
1193238	
2 (bases 1 to 11759)	
Birren,B.	
Direct Submission	
Submitted (20-MAR-2002) Center for Genome Research, Whitehead	
Institute, Nine Cambridge Center, Cambridge, MA 02141, USA	
Location/Qualifiers	
1..11759	
/organism="Methanosarcina acetivorans CZA"	
/strain="CZA"	
/db_xref="taxon:188937"	
complement(1002..1544)	
/gene="MA3789"	
complement(1002..1544)	
/gene="MA3789"	
/codon_start=1	
/transl_table=11	
/product="nitroreductase"	
/protein_id="AA07140.1"	
/db_xref="GI:19917862"	
/translation="MPVTGKRVVDITLIRNSVREFTTRPISKEPITILNAGKRAP	
SLINQPMRFVAVRNPEITLOKLSCEHYGIVAGALLLAELDEHSNRRKDIQAI	
ASIONMLACGSLGAVMLGELINQHKVNSIIGCSSKLEIMAVLAIGEPAPGRFS	
TRKALPELVPEEKYGOXKE"	
complement(1945..2952)	
/gene="MA3790"	
/note="MA3790"	
complement(1945..2952)	
/gene="MA3790"	
/codon_start=1	
/transl_table=11	
/product="ketol-acid reductoisomerase"	
/protein_id="AA07141.1"	
/db_xref="GI:19917863"	
/translation="MATIYDNETTFDALKDRTIATMGSGQGHANHLHESQNLNV	
VGIRKSSMAKESDGLKMTVEEAAAAVIMILDEDOASVYSIPELAEAG	
ALVFRKHGNHYNQIVPRKDYDVEWAKGGHIVRRYTGIGVPIIAYGORAT	
ARDALAAKGIKATRAVYPTTREPETETDLRGQVYDLGSLSLITAEVLEAG	
YQPEAITECHEVALLVDLITEGGLERHMHVSNTAIEGATVGPRTIINESREMY	
EALARIQGEFEKFEVLGWNHPIKMERQEKHPLEVVKEIIRANIPLNKKIDD	
D"	
complement(3314..3799)	
/gene="MA3791"	
complement(3314..3799)	
/gene="MA3791"	
/codon_start=1	
/transl_table=11	
/product="acetylactate synthase, small subunit"	
/protein_id="AA07142.1"	
/db_xref="GI:19917864"	
/translation="MKHTLAVLVENKGSVRSVLSFRRKYNIDSLAVGTEDPDIS	
RITTVAGDDHVLQVTKLIDIVIKVSDIGDDAVRELAIKVSADVNTAEII	
OIANIFRARIIVDAPKSMVTVEGTDESKIQAIKRLRQGIKEMVRTGIAMVGPCK	
V"	

SOURCE
ORGANISM
Guillardia theta.
Chloroplast Guillardia theta
Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
REFERENCE
1 (bases 47701 to 48415)
Douglas,S.E. and Durnford,D.G.
The small subunit of ribulose-1,5-bisphosphate carboxylase is
plastid-encoded in the chlorophyll c-containing alga Cryptomonas
phi
JOURNAL
MEDLINE
Plant Mol. Biol. 13 (1), 13-20 (1989)
93357429
PUBMED
2562756
REFERENCE
2 (bases 18535 to 19351)
Douglas,S.E. and Durnford,D.G.
Sequence analysis of the plastid rDNA spacer region of the
chlorophyll c-containing alga Cryptomonas phi
JOURNAL
MEDLINE
DNA Seq. 1 (1), 55-62 (1990)
92119320
PUBMED
2132959
REFERENCE
3 (bases 43739 to 44938)
Douglas,S.E. and Durnford,D.G.
Nucleotide sequence of the genes for ribosomal protein S4 and
tRNA(Arg) from the chlorophyll c-containing alga Cryptomonas phi
JOURNAL
MEDLINE
Nucleic Acids Res. 18 (7), 1903 (1990)
90245597
PUBMED
2336372
REFERENCE
4 (bases 34539 to 35380)
Reith,M. and Douglas,S.
Localization of beta-phycoerythrin to the thylakoid lumen of
Cryptomonas phi does not involve a signal peptide
JOURNAL
MEDLINE
Plant Mol. Biol. 15 (4), 585-592 (1990)
91338697
PUBMED
2102376
REFERENCE
5 (bases 45872 to 47981)
Douglas,S.E., Durnford,D.G. and Morden,C.W.
Nucleotide sequence of the gene for the large subunit of
ribulose-1,5-bisphosphate carboxylase/oxygenase from the
chlorophyll c-containing alga Cryptomonas F: evidence supporting
the polyphyletic origin of plastids
J. Phycol. 26, 500-508 (1990)
JOURNAL
MEDLINE
6 (bases 110917 to 113854)
Douglas,S.E.
Unusual organization of a ribosomal protein operon in the plastid
genome of Cryptomonas phi: evolutionary considerations
JOURNAL
MEDLINE
Curr. Genet. 19 (4), 289-294 (1991)
91303433
PUBMED
1868578
REFERENCE
7 (bases 40675 to 42376)
Douglas,S.E. and Turner,S.
Molecular evidence for the origin of plastids from a
cyanobacterium-like ancestor
J. Mol. Evol. 33 (3), 267-273 (1991)
92099311
PUBMED
1757997
REFERENCE
8 (bases 96129 to 98906)
Wang,S.L. and Liu,X.Q.
The plastid genome of Cryptomonas phi encodes an hsp70-like
protein, a histone-like protein, and an acyl carrier protein
Proc. Natl. Acad. Sci. U.S.A. 88 (23), 10783-10787 (1991)
JOURNAL
MEDLINE
92073372
PUBMED
1961745
REFERENCE
9 (bases 106789 to 108216)
Douglas,S.E.
A secy homologue is found in the plastid genome of Cryptomonas phi
JOURNAL
MEDLINE
FEBS Lett. 298 (1), 93-96 (1992)
92183838
PUBMED
1544427
REFERENCE
10 (bases 42198 to 44153)
Douglas,S.E. and Reith,M.E.
A bohl homology, encoding a subunit of Mg chelatase, is located on
the plastid genomes of red and cryptomonad algae
J. Mar. Biotechnol. 1, 135-141 (1993)
JOURNAL
MEDLINE
11 (bases 82327 to 84479)
Douglas,S.E. and Murphy,C.A.

TITLE
JOURNAL
REFERENCE
12 (bases 98901 to 114602)
Wang,S.L., Liu,X.Q. and Douglas,S.E.
The large ribosomal protein gene cluster of a cryptomonad plastid:
gene organization, sequence and evolutionary implications
Biochem. Mol. Biol. Int. 41 (5), 1035-1044 (1997)
JOURNAL
MEDLINE
9178335
PUBMED
9283757
REFERENCE
13 (bases 61067 to 68605)
Leitsch,C.E.W., Kowallik,K.V. and Douglas,S.E.
The alpha gene cluster of a cryptomonad, Guillardia theta: A piece
in the puzzle of chloroplast genome development
J. Phycol. (1998) in press
JOURNAL
MEDLINE
14 (bases 1 to 121524)
Douglas,S.E. and Penny,S.L.
The plastid genome of the cryptophyte alga, Guillardia theta:
complete sequence and conserved syteny groups confirm its common
ancestry with red algae
J. Mol. Evol. 48 (2), 236-244 (1999)
9929392
PUBMED
99128221
REFERENCE
15 (bases 1 to 121524)
Douglas,S.E.
Direct Submission
Submitted (08-JAN-1998) Institute for Marine Biosciences, National
Research Council, 1411 Oxford Street, Halifax, Nova Scotia B3H 3Z1,
Canada
On or before Sep 15, 1998 this sequence version replaced gi:11396,
gi:11297, gi:18103, gi:18281, gi:11383, gi:11407, gi:12539,
gi:336730, gi:11300, gi:398949, gi:2661180.
FEATURES
source
1..121524
Location/Qualifiers
/organism="Guillardia theta"
/organelle="plastid:chloroplast"
/db_xref="taxon:55529"
/note="plastid"
complement(5..86)
/gene="crn1(gua)"
complement(5..86)
/gene="crn1(gua)"
/product="rRNA-Tyr"
/note="codon recognized: UAC"
/anticodon=(pos:50..52,aa:Tyr)
complement(181..573)
/gene="rpl19"
complement(181..573)
/gene="rpl19"
complement(181..573)
/codon_start=1
/product="ribosomal protein L19"
/protein_id="AAC3594.1"
/db_xref="GI:3602933"
/translation="MEILHLLEYFYNNKSNKLLQLVESNHNKTLPOLQVQDVTYRLG
VSTIEGKNERIDATGEVLISSKNCINILITIRVLQGIERYVLLHSPMKLSTEIL
RKSRIIRAKIYVLRFRSGKATRLQQRN"
729..3188
/gene="clpc"
729..3188
/gene="clpc"
/codon_start=1
/product="Clp protease ATP binding subunit"
/protein_id="AAC3595.1"
/db_xref="GI:3602934"
/translation="MEERTTEKAKVIMLAQEARLGNFVGTQILLGLIGEGTGI
AAKVLKSMGVNLKDAVEKTIIGSGSFVAVEIPTPRAKRVLELSLEEAROLGHNH
IGTEHLGLIGEGVAVARVLEALDFTKRVTOVIRLGTAVASATNQTCKTP
TLEERGSNLTOKAAGSKIDPVYGRKTEIRVLIQIGRRKNNPLIGSPGKTAIE
GLAQRINRDVPTDIEDKRVYTLIDGLVYAGIKYGEFEERKKTIDIRANVYIV
IDEVHTLIGAGABGATIDANLTKPALRGEMOCIGATTLDEYRHKIKDSALERFO
PVWGEPSVEETITELVGI,RODYERKHNH,LVISDEALSAARFADYIADRLPKDAID
LIDEGSVRLMNSQLPPAARELDKLEIEIKQDEVAASODFEFAGOLRREMEIKA
GTAIAHAKKGDDEMTKEVSVTEEDIAQIVAAATGIVNNKMTRESEKLLQMEETLH

```
GHIGODEAVNAVSKAIRRAVYGLKNPNRIASFJESGPTGVGKTELLTAKALASYEFGS
BEAVNRLDMSEYMERHTVSKLIGSPGYNGEGLJESVRRPYTVLFPDIEGHP
PEAVNLLDLEDGRITDSKGRVDEKNTLLITSVGSKVIEKGGGVLPELSEOT
ESQYGRKALVNEBELKQVRRPELINDDIIVYRQITREVECEIAPIMKEVPTRISE
KGIQLEVTARFETHLINEGYNPIYCARPLRRAVMRLLEDJTLSEEFIAEKIKGDAIVV
DVDDDKRVALLGEEKLELVAN"
/ gene
/ gene="trnf(gaa)"
/ protein_id="AAC35596.1"
/ 3316..3388
/ gene="trnf(gaa)"
/ product="trnA-Phe"
/ note="codon recognized: UUC"
/ anticodon=(pos:3349..3351,aa:Phe)
/ complement(3402..5201)
/ gene="dnaB"
/ complement(3402..5201)
/ gene="dnaB"
/ codon_start=1
/ transl_except=(pos:complement(5199..5201),aa:Met)
/ product="replication helicase subunit"
/ protein_id="AAC35596.1"
/ db_xref="GI:3602935"
/ translation="MKVTKFYMLKAEKHLIIQILVEKDFEIQIYTKIDTQVYFQSHQ
ILYRSINILHTRKVLINLPNLSLHSELVSEVINYLLTFNDRNFRILNIDSLK
VLIDNFIRELONSCAKIVSLTFENMLSEVTELOKSNLIISNITNKKVSLKISOL
LLEITLIDKTNRSTHVLGFEFDLHILVGLQKSLIITACRPSMGKTAFLMILVRN
VDIOSFPVIVISLEMSKQILYRLISNTEINATSRREGNISISPEMLINAMTIS
NINIVYDDENNDDVIDIDIOSKLASLOOITYDIGLIVDIYQILQYKRKSSQRITELSTI
TRKLIADKFNPLPVLSQLSRNVSRNKRPLISDLKESCLSTAIQVPELFSDN
VNESKLTLSKVNHYILSATLELLIPQYNRIYPIVSLIKRELQYGRVYVLEDFYIS
VIVSTEHYVLLNMGKRILLELVDDLVATLIDIOYLIYNTEVDLEFSNVYSVNL
ICMNRINVDPMIPIKTNFNEVALLVHNSIEDODAVLFLYRDSYVQHMLKQSNIDMC
EVIYAKHRGHTGMVLIINPMTVSFNMLIKES"
/ misc_feature
/ note="signal element intein: encodes putative intein
spliced out of the replication helicase subunit"
/ gene="trnW(cca)"
/ 5361..5433
/ gene="trnW(cca)"
/ 5361..5433
/ gene="trnW(cca)"
/ product="trnA-Tyr"
/ note="codon recognized: UGC"
/ anticodon=(pos:5394..5396,aa:Tyr)
/ 5588..6013
/ gene="rpl11"
/ 5588..6013
/ gene="rpl11"
/ codon_start=1
```

Query Match 8.7%; Score 145; DB 8; Length 121524;

Best Local Similarity 56.0%; Pred. No. 7.1e-30;

Matches 274; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

```
QY 999 CATACTCTATCATGCTGTTAAATGATATTCACAGAGCTTCTATATTTGACTGCTGTT 1058
||||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 50000 CATACATTATACAGTTTGTGTTGAAGATGATCTGTTTAACTCGTATGCTGTTTA 49941
QY 1059 TTGGCTCGAAGGATATACATCCAGAGCTTGCCCTAGACATGCTGAACCAAGGCC 1118
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49940 TTGGCTCGCTGTTCAATATCGAAGTTTACAGCACTGGACCACTGAAAAATTAGGA 49881
QY 1119 ATTTCAGCATTCACACAGATTAACCTGCAACAGATGAATCGCTGACGAATTTGCTGAC 1178
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49880 ATTTCAGCATTTACTATGCTGTTCCATCGGATGAAGACATATTAGCAATTAACAAA 49821
QY 1179 CAACCTTACAACCTGATGATGTCATGAGGTCATGATCTTACTCATTTGCCATTTTCT 1238
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49820 CAACCTTATAGCTGTTATATATTAAAGTTGAAGATATTAGCAATCTACCTACGGTT 49761
QY 1239 GAAAGAGAACTGATGCTGATTAAGATTGCCGTGACCGCTGCTGTAAGAGATGTCCTG 1298
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49760 GACCGAGAACTGATGTTGAATTAACGAGTTTCAACATTTAGAGAGACTGAAGCATTA 49701
```

```
QY 1299 GACATTGCTAGTATATTTTCAGGCTAAAGCTGTTGACGTATCTGATCAACAACTACTTG 1358
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49700 GATATATTAATTAATTAATTTGCTGCTAGGTTGCTGACCTATCTGAGATTTTAAATTA 49641
QY 1359 CACCTTACGCGGATCTAGACAGATGTTGACCTGACCTCAAGGTTATTGAGCCCTATGCT 1418
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49640 GAAGTTACAGAGTGATCCCTGTAATAATTTGTCATAGACCAAAATTTTGTCAAAATTTACT 49581
QY 1419 ATATGAGAGCTGACAGAACCCGCTGCTGCGCTTTGGCTGCGTAATCGGAGGAGCTCC 1478
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49580 ATATCTGAGATAGTAAAGACTGAAAAATTTCTTGTACGCGAATCTAATGTTAATACC 49521
QY 1479 AAGTACCTT 1487
| | |
Db 49520 CGTTCTCTT 49512
```

```
RESULT 15
AP003597 342850 bp DNA linear BCT 28-NOV-2001
LOCUS Nostoc sp. PCC 7120 DNA, complete genome, section 17/19.
DEFINITION AP003597 Ba000019
ACCESSION AP003597 GI:17133729
VERSION AP003597.1 GI:17133729
KEYWORDS .
SOURCE Nostoc sp. PCC 7120 DNA.
ORGANISM Nostoc sp. PCC 7120
Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
REFERENCE 1
AUTHORS Kaneko,T., Nakamura,Y., Wolk,C.P., Kuritz,T., Sasamoto,S.,
Watanabe,A., Iriuchij,M., Ishikawa,A., Kawashima,K., Kimura,T.,
Kishida,Y., Kohara,M., Matsumoto,M., Matsuno,A., Muraki,A.,
Nakazaki,N., Shimo,S., Sugimoto,M., Takazawa,M., Yamada,M.,
Yasuda,M. and Tabata,S.
Complete genomic sequence of the filamentous nitrogen-fixing
Cyanobacterium Anabaena sp. strain PCC 7120
DNA Res. 8 (5), 205-213 (2001)
2 (bases 1 to 342850)
Kaneko,T.
Direct Submission
Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research; Yana
1532-3, Kasarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp.
URL:http://www.kazusa.or.jp/cyanobase/,
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
FEATURES
source
1. 342850
/organism="Nostoc sp. PCC 7120"
/db_xref="taxon:103690"
/note="synonym:Anabaena sp. PCC7120"
complement(174..1478)
/ gene="a114593"
/ complement(174..1478)
/ gene="a114593"
complement(174..1478)
/ note="ORF ID:a114593
hypothetical protein"
/ codon_start=1
/ transl_table=11
/ protein_id="BAB76292.1"
/db_xref="GI:17133730"
/ translation="MDKSILOLVDELPSNITVVKVNLADYVAPGOMONLVGFNVTIR
AVTGFDARVIOKIRDRATSLYNDPOCGQSVKRYXRTIDRADVTMAATLANKYGEK
IGFLSFLSITPRKAVYOSIDLVSKAIEIIFYCKLNGIPQNPPEFRANSLANNYQDA
SLIRNVALVCLDGILPLGPDELSTIQGIISGAEVSNITQNVFLVANSFLPGSNFTDK
LGFINQGFNAVQGMNMLVTKGITPOSISHLGNFIQIADNLDVNASLFDQNTNF
EHTGISOVARRLILKAHTLVKEIKQETTOPIDSSSAVKSDTSQYAAASKLEVDEE
EEDWQATIEKVQDDOFPINVIYIGSSHNEMWKGKDIATDVAASSDANGYAVKKVKC
WSDEEYVATATIEKIQGNQYRIHTGTGDSIDENVNDDIS"
```

```
gene
/feature="ORF_ID:asr4594
hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAB76293.1"
/db_xref="GI:17133731"
/translation="MTAIAPOEKSPROVLYLIGEDNNFVVEVPSLPGCISOGKTRREA
LTNIEVLQDKGETLPEYOFINI"
1877..2770
/gene="alr4595"
1877..2770
/gene="alr4595"
/feature="ORF_ID:alr4595
unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB76294.1"
/db_xref="GI:17133732"
/translation="MSKRLVDRLLLFCFTVLGFTPIIHNVQAESAPFWEFINVDA
VQPGDMLVETQKYNFIRDYNNHLSRYIPLDNDRLTEVSFSENGQLLPSTSTSG
NHQLLSWEHQLQSPESHTEFVLYKRYIGGLVNNHVAQVWEALFSDRKAPIKQAOVR
VEPEELLNKIKIFOSGVVADIRVDKTEAVTKQPIERGTGIQVITDRAGTKI
ETSGMSTQPFYERLWMLLGLVFGIPLILLIYSRNGTGTSSYIGDSSIDYIGGGG
DYGGDYGGDYGGGGGGG"
3082..4473
/gene="alr4596"
3082..4473
/gene="alr4596"
/feature="ORF_ID:alr4596"
/codon_start=1
/transl_table=1
/product="malic enzyme"
/protein_id="BAB76295.1"
/db_xref="GI:17133733"
/translation="MADITPNSFSVTLRLQIPNRVGMASITQAIASSGNGLCQIDL
IEQTRQVSLRDLTVDASTEHAETIVQAVKAIPIKILSVDRFTNLRGKISITSR
IPLKSVSDLAAYIPGVGRCYCTAIAODPSOYNTIKONTVAIVTDSAVLGLNLP
HAALPMEGKAMLFKEFGADLPICLATOPEELIQAKNIAPVGGVNHEDIAPR
CFEIEKRLRDELDPVFHDQGHATVATLALFNSLKVOKSLADIRIVINGAGACV
AIALRLKAGAEKIMDCSKGISTTRFDLNEEFKQAFKAGOTLAGVQADYFICV
SARCVLTPENYKSMAKDAIVPAMNPPEIOTPELISKVAVATGRSDPEQINNVA
EPGVFGALDCRAQITTTMYLEASALASLAVNSPDSREHIIPSVDERVYTVAAA
VQRAAREGIARG"
5270..6853
/gene="alr4597"
5270..6853
/gene="alr4597"
/feature="ORF_ID:alr4597"
/codon_start=1
/transl_table=1
/product="type I restriction-modification system DNA
methylase"
/protein_id="BAB76296.1"
/db_xref="GI:17133734"
/translation="MTEKAARKNNSESEFQKLMKAADKLKRNIDAEXKHIVLGLIFL
KYIDSAPEELHOKLLAGEGYAGANPEDRDEYSAENVFVPEARMSYLOQAKOPDI
KTVGLAMEALIEQENAKRLKGLIPKRYGOKLDOKSLGGLIDLGSTITLGAEOAOD
VLGVYVEYELGOFALAEKGKGGOFYTESIVKILVEMLEPYNGSVFDPDCCSGGAFV
SEKPVKNHOGRLDISITGQSNETTYIKLCRMNLAIRGIDCSNITKMNESGFLLDAK
DLKADFYIANPFNDSDWGGELLNRDGRKMDKDLVPPGNAFPAMVSHFIYHLAPTGS
AGFVLISNGSLSSNTEGEGDIRKALVQKDLVCIYWLPIQLFYNTGIPACLWFLPYKN
GKNRNRDRGEVLFIDASELGYNNVSRAPFAEALISKIADTYHEMKKSGGSYRIDKGF
CKNSASIAIEIEKHNEFVLPGRYVGIPDEVEDGVSEPEKMSSELTMALGEGMRGQLDDEE
IKKOLLKYGFIYSELVDVF"
6858..7121
/gene="asr4598"
6858..7121
/gene="asr4598"
/feature="ORF_ID:asr4598
unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB76297.1"
```

```
gene
/db_xref="GI:17133735"
/translation="MNMTOVILKKNLIVIEKLLKHLAQSORTLEETITSLLEDVTEN
TPITTSKSRDMSPGCFQTCAGMGGELLVRPQPEAQRERPL"
7226..7804
/gene="alr4599"
7226..7804
/gene="alr4599"
/feature="ORF_ID:alr4599
hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAB76298.1"
/db_xref="GI:17133736"
/translation="MIAVKGKFPKLTPEEYFIMEKKOQLRHEYLNGEVYAMTGSTONH
GRISNIIFIVIKHGLRGCGGVNSDORINFEFKDYVYPVPSVCDERPTATOAO
YPCILVEVLSPTASTSYDRGKFRILYRRRPSIQDILVDAEKAIADILRKNERGMWEIF
NYOLGDNKIDSLSPFIESVEDIVFEESV"
7855..8145
/gene="asr4600"
7855..8145
/gene="asr4600"
/feature="ORF_ID:asr4600
unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB76299.1"
/db_xref="GI:17133737"
/translation="MIREKNMTOVILKOLNPYLEKIKRAQSORTLEETITSLILE
DVAREEPKSSSEGFMDWTLOFRERMOENITFDADPADIRDSVGRDVEF"
8211..8573
/gene="alr4601"
8211..8573
/gene="alr4601"
/feature="ORF_ID:alr4601
hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAB76300.1"
/db_xref="GI:17133738"
/translation="MVKRLTOHOLETATGVSVMHLLFCGLRLVESOKRRLLEYINO
IPLMTIINTDLKAQWMAOERARLSKGTTPATIDQIASIVSNLITLVNVSDF
ESFNDLAVENFVNSGEG"
8577..9749
/gene="alr4602"
8577..9749
/gene="alr4602"
/feature="ORF_ID:alr4602"
/codon_start=1
/transl_table=1
/product="type I site-specific deoxyribonuclease chain S"
/protein_id="BAB76301.1"
/db_xref="GI:17133739"
/translation="MSEWKEETTLGEIADIINGQSPGTGETCNNNGQLPLNGPTEFGD
RNLPLOTPTIDPKRIAEAGDGLFCVRSSTGRMMWADQKVAIGRGIASIRAKDGLFQ
PYIRALIEKEKLSLAAVATGSTPFINISPDHNLNLIVOLPSKNIKITYISNLRIIDEKI
YNIILSONTEPLAIKOTLEKHFIDPEEPNADGKPKYKSSGAMVMSALCYIDEAMSVEK
LGOVLNIRKHGVAFGVETTEPTEKILITLPNPFITGGGFNSKYYASADYSNEVYL
RRKDLATMDLSKGBSGLGPATIPDILKGVPLIANDRIGKVENNNIDKTLTYLLCR
REYSHITIGTSSGSTVSRHTSPSRICEYSFVLPDELIDKFSALATATIDKILISTL"
9944..10465
/gene="alr4603"
9944..10465
/gene="alr4603"
/feature="ORF_ID:alr4603
unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB76302.1"
/db_xref="GI:17133740"
/translation="MICKLREINDERNFSHIALAESQADKYNNLLISKYIDLFEV
KKLSISILQYKNTISNTIDILFLKFDGHSLSKKRNHDLIVNNFIRTNIDVNEIRLP
CKFIANNQIICLSPAYGVLNGYPIHLFYKKAQKEPNYFIFEVIADQPREIKIERNI
FDVSIQIIESLLL"
```


GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2003, 18:38:17 ; Search time 417 seconds
(without alignments)
9034.996 Million cell updates/sec

Title: US-09-997-900-1
Perfect score: 1673
Sequence: 1 gctcttcagcagcaaaaa.....tgatttggcttcattcg 1673

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_101002.*
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1673	100.0	1673	21	AAA27424
2	875.6	52.3	960	24	ABN98543
3	701.8	41.9	1721	24	ABA98820
4	665.8	39.8	1861	19	AAV55857
5	593.8	35.5	1435	24	ABA98819
6	590	35.3	1813	24	ABA98816
7	537.6	32.1	1287	24	ABA98814
8	419	25.0	4895	21	AAA27425
9	243.4	14.5	611	19	AAV55854

10	201.8	12.1	724	24	ABA98822
11	197	11.8	470	24	ABA98813
12	169.4	10.1	515	24	ABA98817
13	160.8	9.6	484	24	ABA98818
14	149.4	8.9	417	24	ABA98821
15	147.6	8.8	420	21	AAC37251
16	112	6.7	565	24	AAD27618
17	97.8	5.8	547	24	ABA98815
18	92	5.5	2320	19	AAZ96290
19	91	5.4	525	22	AAS00993
20	91	5.4	3291	22	AAS02313
21	91	5.4	10426	22	AAS02329
22	90.4	5.4	6211	19	AAV52141
23	90	5.4	501	23	AAS55637
24	90	5.4	501	23	AAS55660
25	79.8	4.8	13732	21	AAA81484
26	79.8	4.8	349980	21	AAE21611
27	76	4.5	12720	15	AA064211
28	74.4	4.4	516	22	AAH68423
29	74.4	4.4	639	22	AAE71477
30	74.4	4.4	2427	22	AAE54905
31	74.4	4.4	349980	22	AAH68528
32	71	4.2	489	24	AB090233
33	64.8	3.9	640681	24	ABA92787
34	64.6	3.9	522	23	AAS56283
35	62.6	3.7	1489	23	AAS51405
36	62.6	3.7	522	23	AAS54008
37	62.6	3.7	96109	22	AAE28548
38	62.6	3.7	319630	24	AB067194
39	62.6	3.7	3011208	24	AB069245
40	61.6	3.7	2286	24	AB069168
41	61.6	3.7	6506	24	AB071032
42	60	3.6	561	17	AAT45894
43	57.8	3.5	2944528	24	ABA03041
44	56.8	3.4	492	23	AAS54309
45	55.4	3.3	492	22	AAH78452

ALIGNMENTS

RESULT 1	AAA27424	
ID	AAA27424	standard; cDNA; 1673 bp.
XX	AAA27424;	
AC		
XX		
DT	11-AUG-2000	(first entry)
XX		
DE	Arabidopsis acetylhydroxy-acid synthase (AHAS) small subunit cDNA.	
KW	Herbicide-resistance; acetylhydroxy-acid synthase; AHAS;	
KW	acetylacetic synthase; imidazole; sulfonylurea;	
KW	triazolopyrimidine sulfonamide; sulfonylurea; enzyme;	
KW	pyrimidin-oxo-benzoic acid; sulfonylcarboxamide; transgenic plant;	
KW	branched-chain amino acid synthesis; ss.	
OS	Arabidopsis sp.	
XX		
XX		
FT	Key	Location/Qualifiers
FT	CDS	42..1517
FT		/*tag= a
FT		/product= "Arabidopsis AHAS small subunit"
FT		/Ec_number= 4.1.3.18
XX		
XX		
PN	WO200026390-A2.	
XX		
PD	11-MAY-2000.	
XX		
PF	28-OCT-1999;	99MO-US25452.
XX		
XX	29-OCT-1998;	98US-0106239.
XX		

Wheat ALS small su
Corn ALS small sub
Rice ALS small sub
Rice ALS small sub
Wheat ALS small su
Arabidopsis thaliana
Synchocystis sp.
Corn ALS small sub
S. pneumoniae deri
B. subtilis gene e
B. subtilis llyBNC
Plasmid PAN267 for
Streptococcus pneu
Streptococcus pneu
Streptococcus pneu
N. meningitidis pa
Neisseria meningit
L. lactis branched
C glutamicum codin
Corynebacterium q1
Nucleotide sequenc
C glutamicum codin
M. capsulatus gene
Buchnera sp. genom
Salmonella typhi D
Klebsiella pneumonia
Klebsiella pneumonia
Genomic fragment #
Listeria innocua c
Listeria innocua D
Listeria monocytog
Listeria monocytog
R. capsulatus acet
Listeria monocytog
Pseudomonas aerugi
DNA of small subun

PA (AMCY) AMERICAN CYANAMID CO.
XX
PI Kakefuda G, Costello C, Sun M, Hu W,
XX
DR WPI: 2000-365633/31.
XX P-PSDB; AAY96203.
XX
PT New polynucleotide encoding eukaryotic acetoaldehyde-acid synthetase
PT small subunit protein for producing transgenic herbicide resistant
PT plants and identifying mutations affecting enzymatic activity of the
PT synthetase -
XX
PS
PS Claim 2; Page 47-50; 57pp: English.
XX
XX The present sequence is the cDNA for Arabidopsis acetoaldehyde-acid
CC synthase (AASs) small subunit protein. AASs is also known as
CC acetoaldehyde synthase. This enzyme is needed for branched-chain amino
CC acid synthesis and so is essential for life. Inhibition of this enzyme
CC would lead to plant death and therefore inhibitors would be potential
CC herbicides. Certain herbicides are known to inhibit AASs:
CC pyrimidinolones, sulfonylureas, triazopyrimidine sulfonamides,
CC pyrimidyl-ox-benzoyl acids, sulfonylureas and sulfonylcarboxamides.
CC The present sequence may be used to identify mutant AASs which are
CC resistant to these herbicides and may be used to create herbicide
CC resistant transgenic plants e.g. dicot and monocot crop plants.
XX
XX Sequence 1673 BP; 466 A; 328 C; 418 G; 461 T; 0 other;

Db	601	TTGACACATTCAGACCAAGAGTTGTAGATATACGCCAACTGCATTGGTACTATCGAGGTA	660
Oy	661	CTGCGATCCTGGAAAAATGATTGCTGTAGAAAAGAAATTTGAAAAAGTTTCAGATCAGAG	720
Db	661	CTGGAGATCTCGAAAAATGATTGCTGTAGAAAAGAAATTTGAAAAAGTTTCAGATCAGAG	720
Oy	721	AGATTGTAAGCACAGGAAGATAGCACTGTAGAAAGGAAAGATGGGTCAACCTGCTCAT	780
Db	721	AGATTGTAAGCACAGGAAGATAGCACTGTAGAAAGGAAAGATGGGTCAACCTGCTCAT	780
Oy	781	TTTGGCGATTTTCAGCAGCATCTTATCCAGATCTTAAGGAGCAAGCGCTGTATGTTTC	840
Db	781	TTTGGCGATTTTCAGCAGCATCTTATCCAGATCTTAAGGAGCAAGCGCTGTATGTTTC	840
Oy	841	TTTCGAGTAGCAAAAAAGAGGCATTGTCCCTCAAAAGAAAGAAACATCAGCAGGGGAGATG	900
Db	841	TTTCGAGTAGCAAAAAAGAGGCATTGTCCCTCAAAAGAAAGAAACATCAGCAGGGGAGATG	900
Oy	901	TTTATCCCGTTGAGCCATTTTGTGACCCCAAGGTATCATCTGATTCTGCAGCCTCACTGGG	960
Db	901	TTTATCCCGTTGAGCCATTTTGTGACCCCAAGGTATCATCTGATTCTGCAGCCTCACTGGG	960
Oy	961	GACTTCTCACTAGCAAAATACAGATGAGTACGATTCGCTGCATATCTATATCTTTGTTGTA	1020
Db	961	GACTTCTCACTAGCAAAATACAGATGAGTACGATTCGCTGCATATCTATATCTTTGTTGTA	1020

Query Match	100.0%	Score 1673;	DB 21;	Length 1673;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1673; Conservative	0;	Mismatches	0;	Gaps 0;

1021 ATGATATTCAGGAGTCTTAATATTGACTGTGTTTCGTCGAAGGGGATACATA 1080

OY	1	GTCTTCTACATACCAAAAAACCTGGCTGCTCTGTCATGAGGGGCACTTTCGTAA	60
Db	1	GTCTTCTACATACCAAAAAACCTGGCTGCTCTGTCATGAGGGGCACTTTCGTAA	60
OY	61	GTCTTCTACATACCAAAAAACCTGGCTGCTCTGTCATGAGGGGCACTTTCGTAA	120
Db	61	GTCTTCTACATACCAAAAAACCTGGCTGCTCTGTCATGAGGGGCACTTTCGTAA	120
OY	121	TATCTGTAGCCGCTGTAATGCTCTCCGGCGAAGATTCTCATATCTCTCCGCTATATCTGGC	180
Db	121	TATCTGTAGCCGCTGTAATGCTCTCCGGCGAAGATTCTCATATCTCTCCGCTATATCTGGC	180
OY	181	ACCGTGGCGATGAATATGCGTAAGAGATGAGAGATTCGTTAGAAAGCTCGATTCGGAAGA	240
Db	181	ACCGTGGCGATGAATATGCGTAAGAGATGAGAGATTCGTTAGAAAGCTCGATTCGGAAGA	240
OY	241	TCTCTGATGCGCTCTTTTCCGAGCTTCATCTCGCACTCCAAAATGGAAGTGGAGAAC	300
Db	241	TCTCTGATGCGCTCTTTTCCGAGCTTCATCTCGCACTCCAAAATGGAAGTGGAGAAC	300
OY	301	ACACAATTTCGATTTTGTGGAGAGCAAAAGGGAATTAATAGATTGAGAGAGTGT	360
Db	301	ACACAATTTCGATTTTGTGGAGAGCAAAAGGGAATTAATAGATTGAGAGAGTGT	360
OY	361	TTGCAAGAGAGAGATACAAATATTGAGAGTCTTCTGTGGTCTGGAACAGAGCAAGGCTC	420
Db	361	TTGCAAGAGAGAGATACAAATATTGAGAGTCTTCTGTGGTCTGGAACAGAGCAAGGCTC	420
OY	421	TATTCACCATAGTGTCTGTGGAGACTGAAGGCTACTTCAGCAGGTCATCGACCAATCC	480
Db	421	TATTCACCATAGTGTCTGTGGAGACTGAAGGCTACTTCAGCAGGTCATCGACCAATCC	480
OY	481	AGAAAGCTGTGAATGCTCTAAGGTTGAAGATATTCACAGTGAAGCGGCAAGTGGAGCGTG	540
Db	481	AGAAAGCTGTGAATGCTCTAAGGTTGAAGATATTCACAGTGAAGCGGCAAGTGGAGCGTG	540
OY	541	AGCTGATGCTGTGTAAGAGTAATGACATCCAAATCCAGGCGAGAGATCATGTGGCTAG	600
Db	541	AGCTGATGCTGTGTAAGAGTAATGACATCCAAATCCAGGCGAGAGATCATGTGGCTAG	600
OY	601	TTGACACATTCAGAGCAAGAGTTGTAGATATVACGGAGACATGCATTTGACTCATGAGGTTAA	660

Db	1081	TCACAGCGTTGGCCCTAGSACATGCTGAAACCAAGGCACTTTCACGCATTACACAGTTA	1140
OY	1141	TACCTGCAACAGATGAATCGGTCAGCAAAATTGGTCAGCAACTTTACAACCTGTAGATG	1200
Db	1141	TACCGCAACAGATGAATCGGTCAGCAAAATTGGTCAGCAACTTTACAACCTGTAGATG	1200
OY	1201	TGCATGAGCGTCATGATCTTACTCTCAATTTGGCATTTTTCGAAAGAGAACTGATGCGGATTA	1260
Db	1201	TGCATGAGCGTCATGATCTTACTCTCAATTTGGCATTTTTCGAAAGAGAACTGATGCGGATTA	1260
OY	1261	AGATTGCCGTGACACGCTCTGCTAGAGAGATGTCCTGACATTCGTAGTATTTTTCAGGG	1320
Db	1261	AGATTGCCGTGACACGCTCTGCTAGAGAGATGTCCTGACATTCGTAGTATTTTTCAGGG	1320
OY	1321	CTAAAGCTGTGACGTATCTGATCAGACAAATTACTTTGCAGCTTACTGCGGATCTAGACA	1380
Db	1321	CTAAAGCGTTGACCTATCTGATCAGACAAATTACTTTGCAGCTTACTGCGGATCTAGACA	1380
OY	1381	AGATGCTTGCACTGCAAAAGGTTATTTGAGAGCCCTATGATATTTGAGAGGTTGCAAGAACCG	1440
Db	1381	AGATGCTTGCACTGCAAAAGGTTATTTGAGAGCCCTATGATATTTGAGAGGTTGCAAGAACCG	1440
OY	1441	GTCGTGTGGCATTTGGCTCGTGAATCGGAGAGTGAGCATCCAACTACCTTGCGGATACCTCT	1500
Db	1441	GTCGTGTGGCATTTGGCTCGTGAATCGGAGAGTGAGCATCCAACTACCTTGCGGATACCTCT	1500
OY	1501	TTCTTTTAAACAGGCTAAACCGTTGCAGAGTGCATTCATCGAACATCAGAAACCTTTGGAG	1560
Db	1501	TTCTTTTAAACAGGCTAAACCGTTGCAGAGTGCATTCATCGAACATCAGAAACCTTTGGAG	1560
OY	1561	GTAAGAGTTTCATTACAGAGCTATGAACCTCAAGACCTCAAGACAGACAGAGACGCGCGATA	1620
Db	1561	GTAAGAGTTTCATTACAGAGCTATGAACCTCAAGACCTCAAGACAGACAGAGACGCGCGATA	1620
OY	1621	TATGTTTGTGACTTGTATTATGAACAATTAAGCTGATTTTGGGCTTCATTTTCG 1673	
Db	1621	TATGTTTGTGACTTGTATTATGAACAATTAAGCTGATTTTGGGCTTCATTTTCG 1673	

ABN98543/C
ID ABN98543 standard; DNA: 960 BP.
XX
AC ABN98543;
XX
DT 01-AUG-2002 (first entry)
XX
DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 311.
XX
KM Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;
disease; crop; thale cress; tolerance factor; insect; pathogen;
nutrition; ds.
XX
OS Arabidopsis thaliana.
PN US200203281-A1.
PD 21-FEB-2002.
XX
PF 26-JAN-2001; 2001US-0770445.
XX
PR 27-JAN-2000; 2000US-178472P.
XX
PA (GORL/) GORLACH J.
PA (ANY/) AN Y.
PA (HAM/) HAMILTON C M.
PA (PRIC/) PRICE J L.
PA (RAIN/) RAINES T M.
PA (YUY/) YU Y.
PA (RAME/) RAMEAKA J G.
PA (PAGE/) PAGE A.
PA (MATH/) MATHEW A V.
PA (LEDF/) LEDFORD B L.
PA (WOES/) MOESSNER J P.
PA (HAAS/) HAAS W D.
PA (GARC/) GARCIA C A.
PA (KRICK/) KRICKER M.
PA (SLAT/) SLATER T.
PA (DAVI/) DAVIS K R.
PA (ALLE/) ALLEN K.
PA (HOFF/) HOFFMAN N.
PA (HURB/) HURBAN P.
XX
PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
PI Rameka JG, Page A, Mathew AV, Ledford BL, Moessner JP, Haas WD;
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
PI Hurban P;
XX
DR MPI: 2002-400781/43.
XX
PT New Arabidopsis thaliana nucleic acid for identifying homologous genes,
producing compositions that modulate the expression or function of its
encoded protein, and mapping functional regions of protein
XX
PS Claim 1: SEQ ID NO 311; 49pp + Sequence Listing: English.
XX
XX The invention relates to an Arabidopsis thaliana nucleic acid (I)
comprising a sequence capable of hybridising under stringent conditions
to a sequence selected from any one of 999 sequences (ABN98233-ABN99231),
given in the specification or its fragment. A polypeptide (II) encoded by
CC (I), a transgenic plant (III) comprising an exogenous nucleic acid or a
CC genetically modified cell (IV) comprising an exogenous nucleic acid, is
CC useful for screening a candidate agent for its biological effect. (I) is
CC useful in identifying homologous or related genes, in producing
CC compositions that modulate the expression or function of its encoded
CC protein, mapping functional regions of the protein and in studying
CC associated physiological pathways. (I) is also useful for the genetic
CC manipulation of cells, particularly plant cells. (I) is also useful in
CC screening assays of various plant strains to determine the strains that
CC are best capable of withstanding a particular disease or environmental
CC stress. (III) and (IV) are useful for screening of biologically active
CC agents, e.g. fungicides, insecticides, etc., for elucidating biochemical
CC pathways. The screened agents are useful in improved methods of treating

CC crops to prevent or treat disease. (II) are also useful in screening
CC programs to identify agents that mimic or enhance the action of tolerance
CC factors. Such agents are useful in improved methods of treating crops to
CC enhance their tolerance to environmental stress. (I) is also useful
CC for enhancing or inhibiting production of a biosynthetic product in a
CC plant. (III) is useful for identifying other mediators that may induce
CC expression of proteins of interest, for establishing the extent to which
CC any specific insect and/or pathogen is responsible for damage to a
CC particular plant, for identifying other mediators that enhance or induce
CC tolerance to environmental stress, for identifying factors involved in
CC biosynthetic pathways of nutritional, commercial, or medicinal value and
CC for identifying productions of nutritional, commercial or medicinal
CC value. (IV) is useful in the study of genetic functional and regulatory
CC for alteration of the cellular metabolism and for screening compounds
CC that may affect the biological function of the gene or gene products.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=99909770445.
XX

Sequence 960 BP; 279 A; 221 C; 191 G; 269 T; 0 other;

Query Match 52.3%; Score 875.6; DB 24; Length 960;
Best Local Similarity 98.7%; Pred. No. 1.9e-261;
Matches 894; Conservative 0; Mismatches 9; Indels 3; Gaps 1;

QY 768 GCAACGCTCCATTTGGCGATTTTCAGCAGCATCTATCCACATGTCAGAGGACGCG 827
DB 960 GCAACGCTCCATTTGGCGATTTTCAGCAGCATCTATCCACATGTCAGAGGACGCG 901
QY 828 CCGTGTAGTGTCTTGGAAATAGCAAAAAGAGCCATTGTCCTCAAAAGAAACATCA 887
DB 900 CCGTGTAGTGTCTTGGAAATAGCAAAAAGAGCCATTGTCCTCAAAAGAAACATCA 841
QY 888 GCGAGGGGAGATGTTTATCCCGTTGAGCCATTTTGGACCCCAAGGTACATCTATTTC 947
DB 840 GCGAGGGGAGATGTTTATCCCGTTGAGCCATTTTGGACCCCAAGGTACATCTATTTC 781
QY 948 GAGCGTCACGCGGAGCTTCTCACTGACGAAGATGAGTGGACTAGGTCGCTACTCTTA 1007
DB 780 GAGCGTCACGCGGAGCTTCTCACTGACGAAGATGAGTGGACTAGGTCGCTACTCTTA 721
QY 1008 TCATTGCTTGAATGATATATTCAGAGGTTCTTAATATGTCAGTGTGTTTCGCTCGA 1067
DB 720 TCATTGCTTGAATGATATATTCAGAGGTTCTTAATATGTCAGTGTGTTTCGCTCGA 661
QY 1068 AGGGGATACATATCCAGAGCTTGGCGGTAGACATGCTGAACCAAGGCAATTTACGC 1127
DB 660 AGGGGATACATATCCAGAGCTTGGCGGTAGACATGCTGAACCAAGGCAATTTACGC 601
QY 1128 ATTACACAGCTTATACCTGCAACAGATGATGCTGACGAAATTTGGTGCAGCAACTTTC 1187
DB 600 ATTACACAGCTTATACCTGCAACAGATGATGCTGACGAAATTTGGTGCAGCAACTTTC 541
QY 1188 AACCTGTAGATGTGATGAGTGCATGATCTTACTCATTTGGCATTTTGTGAAGAGAA 1247
DB 540 AACCTGTAGATGTGATGAGTGCATGATCTTACTCATTTGGCATTTTGTGAAGAGAA 481
QY 1248 CTGATGCTGATTAAGATTTGCCGTGAACGCTGCTGTAGAAAGAGATGCTGTGACATTGCT 1307
DB 480 CTGATGCTGATTAAGATTTGCCGTGAACGCTGCTGTAGAAAGAGATGCTGTGACATTGCT 421
QY 1308 AGTATTTTCAGGGCTTAAGGTTGTCGATGTCATGATGATGATGATGATGATGATGAT 1367
DB 420 AGTATTTTCAGGGCTTAAGGTTGTCGATGTCATGATGATGATGATGATGATGATGAT 361
QY 1368 GGGGATCTAGACAAGATGTTGCATGTCGAAGGTTATTTGGAGCCCTATATATGTAG 1427
DB 360 GGGGATCTAGACAAGATGTTGCATGTCGAAGGTTATTTGGAGCCCTATATATGTAG 301
QY 1428 GTTGCAGAAACGGTGTGTCGATGCTGCTGATGATGAGGAGTGCATGCAAGTACCTT 1487
DB 300 GTTGCAGAAACGGTGTGTCGATGCTGCTGATGATGAGGAGTGCATGCAAGTACCTT 241

Qy	1488	CGGGATACACTCTCTTCTTTTAACAGGCTAAACCCGTTGCAGAGTGCATCGATCGAACATCA	1547
Db	240	CGTGATCTCTCTTTCCCTTTTAAACAGGCTAAACCCGTTGCAGAGTGCATCGATCGAACATCA	181
Qy	1548	GAACCTTTGGAAAGTAAAGTTTCATTACACAGCTGTATGA--ACCTCAAGACAGACAG	1604
Db	180	GAACCTTTTGAAGGTAAGTTTCATTACACAGCTGTATGAGCCACTCAAGACAGACAG	121
Qy	1605	AGAGACTCGCTCGATATATATGTTTGTGACTTTGTTTATGAACAAATTAGCTGATTTTGGGC	1664
Db	120	AGAGACTCGCTGATATATATGTTTGTGACTTTGTTTATGAACAAATTAGCTGATTTTGGCT	61
Qy	1665	TTTCATT 1670	
Db	60	TCATT 55	
RESULT 3			
ABA98820	standard; DNA; 1721 BP.		
XX	ABA98820;		
XX	18-JUN-2002	(first entry)	
XX	Soybean ALS small subunit (clone sdc2c.pk001.b10) coding sequence.		
XX	ALS: Acetolactate synthase; acetohydroxy synthase;		
XX	amino acid biosynthesis; plant; herbicide; ALS small subunit;		
XX	ALS holoenzyme; crop protection chemical; expressed sequence tag;		
XX	EST; ss.		
XX	Glycine max.		
XX	Key	Location/Qualifiers	
XX	CDS	88..1433	
XX	FT	/*tag= a	
XX	FT	/product= "ALS small subunit"	
XX	FT	/EC_number= "4.1.3.18"	
XX	PN	US2001044939-A1.	
XX	PD	22-NOV-2001.	
XX	PF	08-DEC-2000; 2000US-0732618.	
XX	PR	04-JAN-2000; 2000US-174437P.	
XX	PA	(ABELL) ABELL L M.	
XX	PA	(FALCO) FALCO S C.	
XX	PI	(FAMO) FAMODU O O.	
XX	PI	Abell LM, Falco SC, Famodu OO;	
XX	DR	WPI: 2002-113395/15.	
XX	DR	P-PSDB: ABB08416.	
XX	PT	New small subunits of plant acetolactate synthase and nucleic acids	
XX	PT	encoding them, useful in screening for novel crop protection chemicals	
XX	PT	or potential herbicidal compounds based upon holoenzyme inhibition	
XX	PS	Claim 7; Page 37; 57p; English.	
XX	CC	The invention relates to new isolated polypeptides, which comprises an	
XX	CC	acetolactate synthase (ALS) small subunit. Acetolactate synthase is the	
XX	CC	first committed step in branched chain amino acid biosynthesis in plant	
XX	CC	and bacteria. The polypeptides are useful in aiding the discovery of ne	
XX	CC	herbicides that inhibit plant acetolactate synthase activity. The	
XX	CC	polypeptides are useful for preparing plant ALS holoenzyme, which are	
XX	CC	useful in screening for potential herbicidal compounds based upon	
XX	CC	holoenzyme inhibition. The ALS holoenzyme is also useful in screening	
XX	CC	for novel crop protection chemicals. Particularly, the polypeptides can	
XX	CC	be used as a target to facilitate design and/or identification of	

CC	inhibitors of those enzymes that may be useful as herbicides. Thus, the
CC	polypeptides could be appropriate for new herbicide discovery and
CC	design. The polynucleotides are also useful for producing the
CC	polypeptides and for screening for new herbicides. The current sequence
CC	represents the soybean ALS small subunit (clone sdsc2c.pk001.b10) coding
CC	sequence.
XX	
SO	Sequence 1721 BP; 472 A; 401 C; 391 G; 457 T; 0 other:
Query Match	41.9%; Score 701.8; DB 24; Length 1721;
Best Local Similarity	71.3%; Pred. No. 3,1e-207;
Matches	925; Conservative 0; Mismatches 372; Indels 0; Gaps 0;
OY	285 TCGAAGGTGAGGAACACACAAATTTTCAGTATTTTGTTGGAGACGAAAGCGGAATGTTAAT 344
DB	292 TCAAAAGGTTCGTGACACACACGATTTCCGTTCCGTGCGGAGTACAGCGGAATGATTAAC 351
OY	345 AGGATTTCGACGAGTGTTTGCAAGAGAGATATCAATATTGAAGTCTTGTTGGTCTG 404
DB	352 CGGATTGGCGGGGTTCGTGCGTAGGAGAGATCAACATCGAATTCCTCGCGGTGGCCTC 411
OY	405 AACAGAGCAACAGCGCTATTCACCAATAGTGTCTGGAACTGAAAGGTTACTTGACGCG 464
DB	412 AATGAGCAGACAGGGCCCTCTTCACCAATCGTGTGATGAGGACCGATTAAGTGTGGCCCA 471
OY	465 GTTCATCGAACCACTCCAGAAAGCTCGTTATGTTCTTAAAGTTGAAGATATCTCAAGTGA 524
DB	472 GTCATGAGACACTTCAGAAACCTCGCATATGCTTTAAAGGTAGAGATCTTTGAGGGAA 531
OY	525 CCGCAATGTGACCGGTGACGTGATGCTTTGTAAGTGAATGACATTCAGATTCAGGGCA 584
DB	532 CCACAGGTGGAACGTGAACCTGATGCTCATATAAAGTGACCGGATCCGAACACCATGTG 591
OY	585 GAGATCATGTGGCTGTGTGACATTCAGACAGAGTTGTGATATAGGGAGACATGCA 644
DB	592 GAGTTGAAGTGGTGTGTGACATCTTCAGAGTGAAGATTGTGGATATCTCGGAACATTCG 651
OY	645 TTGATATCGAAGTAACTGGAATCTCTGAAAAAATGATTGTGTAAGAAATTTTGGAAA 704
DB	652 GTGACAATTTAGAGTAACTGAGATCCAGGGAAGATGGCCGGTTCAAGAAATTTCCGC 711
OY	705 AAGTTTCAGATCAGAGAGATTGTAAAGACAGAAAGATAGCACTGAGAAAGGAAAAAGATG 764
DB	712 AAGTTTGAATTAAGAAATATAGCCAGAACTGAAAGATTGATTAAGAAAGGAAAAAGATG 771
OY	765 GGTCGAACGTGCCATTTTGGGATTTTCAGAGCATCTATTCAGAGATCTCAAGAGCAA 824
DB	772 GGTGATCTGCTCCATTTTGGGGATTTTCAGGTCTTTATTCACGATCTTAAAGGAAGA 831
OY	825 GCGCCTGTTAGTGTCTTCGAGTAGCAAAAGAGAGCCATTGCTCCCTCAAAAGAAACA 884
DB	832 ACACCTGTTAATAGCCCTTGTGGAGACAAAAATATGAACCTGTGGCCAACTTGATACA 891
OY	885 TCAGCAGGGGAGAGATTATCCCGTTGACGCCATTTTGGACCCCAAGGTACATCGTATT 944
DB	892 CCTGGGGGGAGAGATTATCTCTATAGAACCATGATGATTTTCACTCAATCAAGTT 951
OY	945 CTCGACGCTCACTGGGGACTTCTCACTAGCAAGATAGCAGTGGACTACGGTCCGATACT 1004
DB	952 CTTTGACGCTCACTGGGGGTCTCTCATATGATGAAGATTACCAAGTGGAAATTCGATCACACAG 1011
OY	1005 CTATCATGCTGTGTAATGATATTCACAGAGTCTCTTAATATATGTCAGTGGTGTTCGGT 1064
DB	1012 TTATCTCATGCTGTGTAACGATGCTCTCTGAGTCTTAAACATTTGTTACAGAGATTTTGGT 1071
OY	1065 CGAAGGGATACAAATATCCAGACTTTGGCCGTAGGACATGCTGAACCAAGGCGCATTTCA 1124
DB	1072 AGAAGAGGGTATAACATTCAGAGTTTAGCTGTAAGCAATGCAAGATTGAAGAGCATTTCT 1131
OY	1125 CGCATTTACACAGTTATACCTGCAACAGATGATGGCTACGCAAAATTGGTGCAGCACTTT 1184
DB	1132 CGACTTTAAACATGTGTGTTCTGGGACAGATGATCAATTAGCAAGTGTGGTCCAGCAACTTC 1191

Qy	1067	AAAGGATACAAATATCCAGAGCTTGGCCCTAGACATGCTGMAAACCAAGGCATTTCAGC	1126
Db	1020	ACGAGGGATATACATCCAAAGTTTACTGTGTGACATGCTGAAGTTGAGGGGCTTTCGC	1079
Qy	1127	CATTACACAGTTTATACCTCCCAACAGATGAATTCGTACACAATTTGGTGACACAATTTA	1186
Db	1080	TATTACAACGGTTCTTCCTGGCAGACAGTAGTCAGTTAGTACCAAGTTGGTGACACAATTTA	1139
Qy	1187	CAAACTCGTAGATGTGCATAGTCCATGATCTTACTCATTTTGGCAATTTCTGAAGAAGA	1246
Db	1140	TAAATTTGGTTGATATTTTCATAGAGTTGGGGATATTACTCACTCCATTTGGGGAAAGAGA	1199
Qy	1247	ACTGATGCTGATTAGATTTCCGCTGGAACGGTGGCTGCTAAGAGATTTGCTGGACATTGC	1306
Db	1200	ACTATATGTGATTAAGATTTCTGTGAATGCTGACGGCCGCCCAATGTTCTTGGACATTGC	1259
Qy	1307	TAGTATTTTTCAGGGCTTAAAGCTTTGACGTATTCGATATCACAATAATTACTTTGCAAGTTC	1366
Db	1260	CAGCATTTTGTAGACCAAAAGCTGTGATGTGTGTACCCACATATTAACTCTTGAGACTTAC	1319
Qy	1367	TGGGATCTAGACAAGATGGTGTGCATCGAAGAAGCTATTGGAGCCCTATGGATATATGGA	1426
Db	1320	AGGAAATTTGCATAGATAGTGTGCTTTGGCAGCGGGCTACTAGAGCCTTATAGATTTTGGA	1379
Qy	1427	GGTTGCAGAAGACCGTGTGTGGCATTTGGCTGTGAATGGGAGTGGACTCCAACTACTT	1486
Db	1380	GGTAGCGGAA-CAGACGTCTGGCAGCTGGTAGCTGATAGAGTTGTGGAATTCGAAGTACTT	1438
Qy	1487	TCTGTGATACTCCTTTTCTTTT	1507
Db	1439	GCAGGATATTCATACCCCTTT	1459

RESULT 5	
ABA98819	
ID	ABA98819 standard; DNA; 1435 BP.
XX	
AC	ABA98819;
XX	
DT	18-JUN-2002 (first entry)
XX	
DE	Rice ALS small subunit (clone r10n.pk117.a16) coding sequence.
XX	
KW	ALS: Acetolactate synthase; acetohydroxy synthase;
KW	amino acid biosynthesis; plant; herbicide; ALS small subunit;
KW	ALS holoenzyme; crop protection chemical; expressed sequence tag
XX	EST; ss.
XX	
OS	Oryza sativa.
XX	
Key	Location/Qualifiers
FT	45..1142
FT	CDS
FT	/*tag= a
FT	/product= "ALS small subunit"
FT	/EC_number= "4.1.3.18"
XX	
PN	US2001044939-A1.
XX	
PD	22-NOV-2001.
XX	
PF	08-DEC-2000; 2000US-0732618.
XX	
APR	04-JAN-2000; 2000US-174437P.
XX	
PA	(ABEL/) ABELL L M.
PA	(FALC/) FALCO S C.
PA	(FAMO/) FAMODU O O.
XX	
PI	Abell LM, Falco SC, Famodu OO;
XX	
DR	WPI; 2002-113395/15.
DR	P-PSDB; ABB08415.
XX	

PT New small subunits of plant acetoacetate synthase and nucleic acids
 PT encoding them, useful in screening for novel crop protection chemicals
 or potential herbicidal compounds based upon holoenzyme inhibition
 XX
 PS
 PS
 XX
 XX
 Claim 7; Page 35; 57pp; English.
 CC The invention relates to new isolated polypeptides, which comprises an
 CC acetoacetate synthase (ALS) small subunit. Acetoacetate synthase is the
 CC first committed step in branched chain amino acid biosynthesis in plants
 CC and bacteria. The polypeptides are useful in aiding the discovery of new
 CC herbicides that inhibit plant acetoacetate synthase activity. The
 CC polypeptides are useful for preparing plant ALS holoenzyme, which are
 CC useful in screening for potential herbicidal compounds based upon
 CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
 CC for novel crop protection chemicals. Particularly, the polypeptides can
 CC be used as a target to facilitate design and/or identification of
 CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
 CC polypeptides could be appropriate for new herbicide discovery and
 CC design. The polynucleotides are also useful for producing the
 CC polypeptides and for screening for new herbicides. The current sequence
 CC represents the rice ALS small subunit (clone r10n.pk117.a16) coding
 CC sequence.

SQ Sequence 1435 BP; 413 A; 282 C; 332 G; 408 T; 0 other;

Query Match	35.58;	Score 593.8;	DB 24;	Length 1435;
-------------	--------	--------------	--------	--------------

Method	Score	DD	2D
Best Local Similarity	71.38;	Pred. NO. 1.1e-173;	

Matches 813; Conservative 0; Mismatches 322; Indels 6; Gaps 2;

OY 394 CTGTTGGTCTCAGACAGACAAGGCTATTCCACCATTAGTTGTGTGGAACTGAAAAAGG 455
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dd 19 CCGTGGGCTCACACAGSAGACAAAGGCCATGTCACCATTTGTCTGCCGCACGSACAGGG 78

Db

454 79	TACCTCAGCAGGTCATCGAGCAACTCCGGAAGCGTTAAATGTCAAAGTGTGAATATA TGCTCAACCAGACTCATCGAGCAAGCTTAACAAGCTTTGAATGTGGAAATC
-----------	---

Qy 514 TCTCAGTGAAGCCGCAAGTGGAGCGGTGAGCTGATGCTTGTAAAGTGAATGCATCCAG 573

Db 139 TATCTAAGGAGCCACAGCTTGAAGAGAGAGCTGATGCTTATATAAAATTTAATGTGGAACCCAG 198

QY 574 AATCCAGGCGCAGAGATCATGTGGCTAGTTGACACATTTCAGACAGCAAGTGTGATATAG 6333
| | | | | | | | | | | | | | | | | | | | | |
db 199 ATCAGCGCTCTGAGGCATATGTTTTACTTGTCATATTTTCGAGCGAAGTGTGTGATATTT 258

634 CGAACATGCATTGACATCGAGGTAACCTGACATCCTGAAAAATGATTGCTGAGAAA 6933

694 GAAATTTGAAAAGCTTCAGATCAGAGAGATTGTAAGGACAGGAAGATTAGCACTGAGAA 753

754 GGGAAAGATGGGTGCACCTGCTCCATTTTGGCGATTTTCAGCAGCATCTTATCCAGATC 813

Db 3/9 GTGAATAAATTTGGAGCAACTGCCCGCTTCTGGGATTTTCTGCTGCTCTTACCCAGATC 438

Db 439 TCATATGAGGCATTGCCCAAAAATCTCTTCTTACTCTCTGAATTAAGACAGTCATAGGAA 498

QY 871 CTCAAAAGGAACATC--AGCAGGGGAGATGTTTATCCCGGTGAGCCATTTTTTGACC 927

Db 499 GTTTTGATCAACCATCCATGCTGGGGGCGATGTCATCTCTGTGGAACCTTATGAGGGTT 558

Db 559 CATCCATGAACCAAGTACTGATGCTCACTGGGCGTCCCTTGATGATGAGATTCAAGTG 618

Db 619 GACITTCGATCATCACTACTATCCATCTTGTCAATGATGGCCCTGGTGTCTCAACATTG 678

OY	1048	TGACTGCTGTTTTCGCTCCGAAGGGCATTAATAATATCCAAAGCTGTGGCCGTAGGACATCGTG	1107
Db	679	TTTAAAGGGGCTTTGCTCGCAGAGGCTACAAATAATACAGAGCTTTCTCTGTAGGCCCAAGCTG	738
OY	1108	AAACCAAGGGCATTTTCACGCATTACAAACAGTATAACCTGCACACAGATGAATCGTCAACA	1167
Db	739	AAAAGTCAGGCGCTTTGGCGTATTACAAACAGTTGCTCTCGGAACAAATTAATCCATTGAGA	798
OY	1168	AATTGTGTGACGACCACTTTACAAACTGCTAGATGTGCTAGTGCATGATCTTACTATT	1227
Db	799	AGTTAGTTACGACGACTTTAACAAACTGTTGATGTGCTAGGCTTAAAGTTAATACCTACT	858
OY	1228	TGCCATTTTTCGAAAGAGACATGATCTGATTAAGATGCCGTAAGCGGTGCTGTAA	1287
Db	859	TGCTTTTTCGTAAGAGGACATTAATCTTATCAAGTTCTCTGAAACACTGCTGCTGGA	918
OY	1288	GAGATTCCTGAGACATTGCTAGTATTTTCAGGCGTAAACCTGTTGACGATCTGATCACA	1347
Db	919	GAGACATCTACAGATTAATTTCTGGAATCTTCGGGCAAAATCTGTATGTTTCTGATCACA	978
OY	1348	CATTACTTTGACGCTACTCGGGGATCTAGACAGATGTTGCACGCAAGGATTAATGG	1407
Db	979	CTGTAGCTTACGACCTTACTCGGGGATCTGCACAAATGTTTGATTTACAAAGCTGTGG	1038
OY	1408	AGCCCTATGCTATATGTGAGCTTGCAAGAACCGGTCGTGSCATTTGCTCGTGANTCGG	1467
Db	1039	AGCCTTATGCGATCTGTGAGGTGCGCCAGAAACAGGGGAGATGGCGGTGATCCCGCATATCG	1098
OY	1468	GAGTGGACTCAANTACTCTCGTGATATACCTTTTCTTAAACAGGCTAAACCGTATGAG	1527
Db	1099	GTGTGCAATCCAACTACTCTGTGCGCTACTCTTTCGTTGTAAATCCCAAGTCTGTGAG	1158
OY	1528	A	1528
Db	1159	A	1159

XX	RESULT 6
XX	ABA98816
ID	ABA98816 standard; DNA; 1813 BP.
XX	
AC	ABA98816;
XX	
DT	18-JUN-2002 (first entry)
XX	
DE	Corn ALS small subunit (clone p0094.cssLI72ra) coding sequence.
XX	
KW	ALS: Acetolactate synthase; acetylhydroxyl synthase;
KW	amino acid biosynthesis; plant; herbicide; ALS small subunit;
KW	ALS holoenzyme; crop protection chemical; ss.
XX	
OS	Zea mays.
XX	
XX	
Key	Location/Qualifiers
FT	1..1452
FT	/*tag= a
FT	/partial
FT	/product= "ALS small subunit"
FT	/EC-number= "4.1.3.18"
FT	/note= "no start codon present"
XX	
PN	US2001044939-A1.
XX	
PD	22-NOV-2001.
XX	
PF	08-DEC-2000; 2000US-0732618.
XX	
PR	04-JAN-2000; 2000US-174437P.
XX	
PA	(ABELL/) ABELL L M.
PA	(FALC/) FALCO S C.
PA	(FAMO/) FAMODU O O.
XX	

F1 Abell LM, Falco SC, Famodu OO;
 XX
 DR WPI: 2002-113395/15.
 XX P-PSDB: ABB08412.
 XX
 PT New small subunits of plant acetylactate synthase and nucleic acids
 XX encoding them, useful in screening for novel crop protection chemicals
 XX or potential herbicidal compounds based upon holoenzyme inhibition -
 XX
 XX Claim 7; Page 31: 57bp: English.
 XX
 CC The invention relates to new isolated polypeptides, which comprises an
 CC acetylactate synthase (ALS) small subunit. Acetylactate synthase is the
 CC first committed step in branched chain amino acid biosynthesis in plants
 CC and bacteria. The polypeptides are useful in aiding the discovery of new
 CC herbicides that inhibit plant acetylactate synthase activity. The
 CC polypeptides are useful for preparing plant ALS holoenzyme, which are
 CC useful in screening for potential herbicidal compounds based upon
 CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
 CC for novel crop protection chemicals. Particularly, the polypeptides can
 CC be used as a target to facilitate design and/or identification of
 CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
 CC polypeptides could be appropriate for new herbicide discovery and
 CC design. The polynucleotides are also useful for producing the
 CC polypeptides and for screening for new herbicides. The current sequence
 CC represents the corn ALS small subunit (clone p0094.cssl121a) coding
 CC sequence.
 XX
 XX Sequence 1813 BP; 545 A; 370 C; 432 G; 466 T; 0 other;

Query Match	35.3%	Score 590;	DB 24;	Length 1813;
Best Local Similarity	68.2%;	Prod. No. 1.9e-172;		
Matches 834;	Conservative 0;	Mismatches 385;	Indels 3;	Gaps 1;
QY 285	TGCAAGGTGAGGAGACACACCAATTTTCAGATTGTTGTGGACGAGAAAGGGAATGATTAT	344		
DB 223	TGCGTTATGAAACCGTCACACGCTATACGTTTTTTGTTGGATGAGAAAGTGGGATGATCAAT	282		
QY 345	AGGATTGCAAGAGAGTCTTTTGCACAGGAGAGATACAAATATTGAGAGCTTGCTTGGCTCG	404		
DB 283	CGAATTTGGCGGGTTTTTGTCTAGAAAGAGGATTAACATGAGATCAATTGGCTGTGGGTTG	342		
QY 405	AACAGACACAGGCTCTATTCCACCTAGTGTCTGTGCACTGAAAGGTAAGTCAACAG	464		
DB 343	AACAAGGATTAAGCACTATTATTACAATAGTAGTGTCAGGACAGACAAAATATTAAACCG	402		
QY 465	GTCAATCGAGCAACCTCCAGAGAGCTGCTAATGTCTTAAAGGTGAAGATATCCAAAGTGG	524		
DB 403	GTCTGATGAGCAACCTAAACCAAGCTTTTAAATGTCATTAAGGTTGATGATTTATCAATTGGA	462		
QY 525	CCGCAAGTGGAGCGGTGAGCTGATGCTTTGTAANAAGTGAATGCACATCCAGAAATCCAGGGCA	584		
DB 463	CCACAAAGTTGAAGAAGAACTTATGCTTTATAAAGTAATGCGAGCGAGAAAAGTTACTT	522		
QY 585	GAGATCATGTGGCTAGTTGACACATTTCCAGAGCAAGAAGTTGTGATATAGCGGAACATGCA	644		
DB 523	GAGATTAAGTGGGTTTGCTTCGCAATTTTCAAAGGAGAGTGGTGAATCTTTCACACTACACA	582		
QY 645	TTGACTATTCGAGTAAGTACGAGATCCTGGAAAATGATGCTGTGTAAGAAATATTGGAAA	704		
DB 583	CTAACTATTGGAGTAACTGTGGAGATCTCTGAAAGATGTTTGCAATACAGAAAGCTCTGAGC	642		
QY 705	AAGTTTCAGATCAGAGAGATTGTAAGGACAGGAAAGATAGCACTGAGAGAGGAGAAAAGATG	764		
DB 643	AAATATGGGATCAGAGAAATTTCTAGAACTGSCAAGATAGCTTTGCGCGCGTGAATAAATG	702		
QY 765	GCTGCAACTGCTCCATTTTGGGCGATTTTCAGAGACATCTATCCAGATCTTCAAGAGCA	824		
DB 703	GGAGAAAGTCCGCTCATTTTGGAGGTTCTCTGAGCTTTTATCCGAGATTTCCAAATGGCA	762		
QY 825	GGCGCTGTAGTCTTCTCGAGTAGTGAACAAAAGAGAGCCATGTGCTCCCTCAAAAGAACCA	884		
DB 763	ATACCTTCAAAATTTTCCAGCAAAATCTGGTGTGCAAGCAATTAATATAGAAATCCGATAGGA	822		

QY	885	TCACAGGGGAGATGTTTAAATCCGTTGAGCCATTT---TTTGACCCCAAGGTACATCGT	941
Db	823	TCCTTCAGGGGGTGTATGTTTATCCAGTGGATCTTTATGAAGCTTCTCATCAACAGTCAAA	882
QY	942	ATTCTCAGCGCTCACTGSGGGACTTCTCAGTCAGCAAGATACGAGTACAGTACGGTCAT	1001
Db	883	ATTCTCGATGCTCATCTGTGGGTGTTATGACTGATGCGGATCCAAACAGGGTTTGTTCACAT	942
QY	1002	ACTGATCATTTGCTGTGAATGATATTTCCAGGAGTTCTTAATATTGTGACTGCTGTTTC	1061
Db	943	ACTGTATCAATCTCTTGGAATGATGTCCTCGAGTCTCAATCTTGTAACAGTGTATTC	1002
QY	1062	GCTGGAAGGGATPACATTAATCCAGAGCTGGCCGTGTGACATCGTAACCAAGGACAT	1121
Db	1003	TTCCAGAAAGGGGCTACATATTCCAGAGCTTTGCTGTTGGCCCACTGTAAACAGAGAACT	1062
QY	1122	TCACGCATTACACAGATTATACCTGCAACAGATGAATCGGTGACGCAAAATGGTCAGCA	1181
Db	1063	TCCTGCAATCACTACTGTTGTTCTCGAACAGATGAATTCATTGCCAAGGTAGTACATCAA	1122
QY	1182	CTTTACAAATCCGTAAGATGTGCATGAGGTCATGATCTTACTCATTTGGCATTTTCTGAA	1241
Db	1123	CTGTACAGCTCATTTGATGTTTATGAAGTTCAGAGATTTTACCCACTTACCATTTGCTGCT	1182
QY	1242	AGAAACATGATGCTGATTAGATTGCGCGTAACAGCTGCTGTAGAAAGAGATGCTCTGAC	1301
Db	1183	AGAAAGTATTATGATCAATAAAGTCCGGGCAAAATGCTACAGCTCGAAGGATGTCTTAGAT	1242
QY	1302	ATTGCTAGTATTTTCAGCGGCTAAAGCTGTTGACGATCTGATCACACAATTACTTTCGAG	1361
Db	1243	ATTGCTCAGATTTTTGTGAGCAGCAAAAGTTGACATPACTAGACACACAAATTACACTAGT	1302
QY	1362	CTTACTGGGATCTACGACAGATGGTTGCACATGCAAAAGTTATTGGAGCCCTATGTGTATA	1421
Db	1303	CTCACCGGAGACATGTGACAAATGGTTTAATTGCAAAAGATCTGACAGCAGATGCGATC	1362
QY	1422	TGTGAGTTTCAAGAACCCGTCGCTGTGGCATTTGGCTCGTGATCCGGAGTGGATCCCAAG	1481
Db	1363	TGTGAGTTTCCAGCAGACAGCCCGGATTTGCTCTGCTCCGAGAGTCTGGAGTTGACTCCAG	1422
QY	1482	TACCTTCGTGAGTACTCTTTC 1503	
Db	1423	TACCTCCGCGGTTTCCCTCC 1444	
RESULT 7			
ABA98814			
XX	AC	ABA98814 standard; DNA: 1297 BP.	
XX	ABA98814:		
XX	DT	18-JUN-2002 (first entry)	
XX	XX	Corn ALS small subunit (clone cen3n.pk0112.c11) coding sequence.	
DE	XX	ALS: Acetolactate synthase; acetylhydroxy synthase;	
KM	XX	amino acid biosynthesis; plant; herbicide; ALS small subunit;	
KW	XX	ALS holoenzyme; crop protection chemical; ss.	
OS	XX	Zea mays.	
XX	XX		
FH	Key	Location/Qualifiers	
FT	CDS	2..1054	
FT		/*tag= a	
FT		/partial	
FT		/product= "ALS small subunit"	
FT		/EC_number= "4.1.3.18"	
XX		/note= "no start codon present"	
FN	US2001044939-A1.		
DD	22-NOV-2001		

XX	08-DEC-2000; 2000US-0732618.
PF	
XX	
PR	04-JAN-2000; 2000US-174437P.
XX	
PA	(ABELL/) ABELL L M.
PA	(FALCO) FALCO S C.
PA	(FAMO/) FAMODU O O.
PI	
Abell LM,	Falco SC, Famodu OO;
XX	
DR	WPI: 2002-113395/15.
P-PSDB:	ABB08410.
XX	
PT	New small subunits of plant acetolactate synthase and nucleic acids
PT	encoding them, useful in screening for novel crop protection chemicals
PT	or potential herbicidal compounds based upon holoenzyme inhibition -
XX	
PS	Claim 7; Page 28; 57Pp; English.
XX	
CC	The invention relates to new isolated polypeptides, which comprises an
CC	acetolactate synthase (ALS) small subunit. Acetolactate synthase is the
CC	first committed step in branched chain amino acid biosynthesis in plants
CC	and bacteria. The polypeptides are useful in aiding the discovery of new
CC	herbicides that inhibit plant acetolactate synthase activity. The
CC	polypeptides are useful for preparing plant ALS holoenzyme, which are
CC	useful in screening for potential herbicidal compounds based upon
CC	holoenzyme inhibition. The ALS holoenzyme is also useful in screening
CC	for novel crop protection chemicals. Particularly, the polypeptides can
CC	be used as a target to facilitate design and/or identification of
CC	inhibitors of those enzymes that may be useful as herbicides. Thus, the
CC	polypeptides could be appropriate for new herbicide discovery and
CC	design. The polynucleotides are also useful for producing the
CC	polypeptides and for screening for new herbicides. The current sequence
CC	represents the corn ALS small subunit (clone cen3n.pk012.c11) coding
CC	sequence.
XX	
SQ	Sequence 1297 BP; 356 A; 263 C; 303 G; 375 T; 0 other;
	Query Match 32.1%; Score 537.6; DB 24; Length 1297;
	Best Local Similarity 69.6%; Pred. No. 3.1e-156;
	Matches 729; Conservative 0; Mismatches 319; Indels 0; Gaps 0.
OY	460 AGCAGCTCATCGACGACAATCCAGAAGCTGGTTAATGTTCTAAAGGTGAAGATTCTCAA 519
Db	3 ACCAAGTCATTGAGCAGCACTCAATAAGCTGCACMGCTTCATGAGTGTGAAGATCTATCTA 62
OY	520 GTGAGCCCCAAGTGAGGGGTGAGGTGATGCTGTTAAAGTGAATCCACATCCAGAAATCCA 579
Db	63 AAGACTCCCTGAGGTTGAANAAGAGAGCTGATGCTTATTAAGCTTAACGTTGACCTGATCACG 122
OY	580 GGGCAGAGATCATGCTGCTAGTGTGACACATTCAGAGCAAGAGTTGTAGATATAGCGGAAC 639
Db	123 GCCCTGAGAGTAGGTTTGTAGTTCACATTTTTTTCAGAGCAAAAATGTTGTATATCTGAGA 182
OY	640 ATGCATTGTACATATGAGGTTAACTGGAAGATCCTGGA AAAATGATTTCTGTAGAACAAATTT 699
Db	183 AAACACTTACCATGAGGTTAGCTGGAGATCCTGGCAAAAATTTGCTCAGTGCAGAGGAATTC 242
OY	700 TGAAAAAAGTTTCACATGTCAGAGAGATGTTGAAGAGACAGGAAGATACACTGACAAGAGAA 759
Db	243 TAAGSAAATTTGGGCATCAAAAGAAATTTTCAGSAGACAGGAAAAATTTGCTTTAGAGACTGAAA 302
OY	760 AGATGGGTGCAACTGCTCCATATTTTGGCGATTTTTCAGCAGCATCTATCCAGATCTCAAGG 819
Db	303 AGATTGGGTGCAACAGCCCGTTTCTTGCGCAATTTTTCGCTGCTCTTATATCAGACCTTATAG 362
OY	820 AGCAAGGCCCTGTAGTGTCTTGTCGAAGTAGCAAAAAGAGAGCCATTGTCCTCCATAAAGG 879
Db	363 AGGCTTTACCAAAAAGAACCGCTTACATCTGTAAATTAAGACAGTGAATGGCAGTTTGTTC 422
OY	880 AAACATCAGCAGGGGGAAGATGTTTATCCCCTGTGAGACCATTTTTTTTACCCCAAGGTACATC 939

```
Db 423 GACCAATCCATGCTGGGGGTGATGTTTATCTGTGGAAATCTTACGAGAGCTCAGCTAAC 482
Oy 940 GATTCTCGACCTCATCGGAGGACTTCTCATGACAAATACAGATGAGTACGGTCC 999
Db 483 AAGTACTTGATCTCATCTGGGCTGTTCTTGATGATGATGATATGCAACTGACTTGTCTCC 542
Oy 1000 ATACTCTATCATTTGCTTGTAAATGATATTCAGAGAGTCTTAAATTTGACTGCTGTTT 1059
Db 543 ATACCTCTCCATCTCTTGATGATGATGCTGCTGTTCTCAACATTTGTAACAGAGTCT 602
Oy 1060 TGGCTCGAAGGGGATACATATCCAGAGCTTGGCCGTAGACATGCTGAACCAAGGCA 1119
Db 603 TTGCTCGAGGGGCTACATATACAGAGCTTGTGCTGCTCAGCTGAGAGGAGGCA 662
Oy 1120 TTTCACGACTTACACAGATTATACCTGCAACAGATGATCGGTGCAACAATTGGTGACG 1179
Db 663 TTTCACGATTATACACAGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 722
Oy 1180 AACTTTACAACTCGTAGATGTCATGAGGTCATGATCTTACTCATTTTGCATTTTCTG 1239
Db 723 AACTTTACAACTCGTAGATGTCATGAGGTCATGATCTTACTCATTTTGCATTTTCTG 782
Oy 1240 AAGAGAACTGATGCTGATTAAGATGCGGTGACAGCTCTGCTGATAGAGATGCTCTG 1299
Db 783 AAGAGAGCTGATGCTTATTAAGGTTTCTGTAACACATGCTGCTGAGGAGAAATTTCTAG 842
Oy 1300 AACTTGCTAGTATTTTTCAGAGGCTAAAGCTGTGAGCTATGTCACACATTTACTTGC 1359
Db 843 AATTTGCTGTAATCTTCTCCAGCAAAACCTATTGACCTTCTGACCTACAGTATACCTTTC 902
Oy 1360 AACTTACTGAGGATCTGACAGAGATGTTGCACTGCAAAAGTTATTTGAGCCCTATGATA 1419
Db 903 AACTTACTGAGATCTGACAGAGATGTTGCACTGCAAAAGTTATTTGAGCCATATGCA 962
Oy 1420 TATGAGAGTTCAGAGACCGGTCGTGTCATTTGGCTGCTGATCGGAGTGGACTCCA 1479
Db 963 TCTGGAGGTCCTCCGCAAGACTGGACAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1022
Oy 1480 AATACCTGCTGATACCTCTTCTTTT 1507
Db 1023 AATACCTGCTGCTGCTTACTCTTCTTCAAT 1050

RESULT 8
AAA27425
ID AAA27425 standard: DNA; 4895 BP.
XX
AC AAA27425:
DF 11-AUG-2000 (first entry)
XX
DE Arabidopsis acetylhydroxy-acid synthase (AHAS) small subunit gene.
XX
KW Herbicide-resistance; acetylhydroxy-acid synthase; AHAS;
KW acetylactate synthase; imidazolinone; sulfonylurea;
KW triazolo-pyrimidine sulfonamide; sulfamoylurea; enzyme;
KW pyrimidyl-oxy-benzoic acid; sulfonylcarboxamide; transgenic plant;
KW branched-chain amino acid synthesis; ds.
XX
OS Arabidopsis sp.
XX
FH Key Location/Qualifiers
FH 1..757
FH promoter /tag= a
FH CDS 758..4739
FH /tag= b
FH /product= "Arabidopsis AHAS small subunit"
FH /EC_number= 4.1.3.18
FH /note= "contains 11 introns"
FH exon /tag= c
FH /tag= 1006
FH /tag= 758..1006
FH /tag= 1007..1084
FH intron
```

```
FT /tag= d
FT /number= 1
FT 1085..1300
FT /tag= e
FT /number= 2
FT 1301..1455
FT /tag= f
FT /number= 2
FT 1456..1534
FT /tag= g
FT /number= 3
FT 1535..1659
FT /tag= h
FT /number= 3
FT 1660..1731
FT /tag= i
FT /number= 4
FT 1732..2236
FT /tag= j
FT /number= 4
FT 2237..2320
FT /tag= k
FT /number= 5
FT 2321..2486
FT /tag= l
FT /number= 5
FT 2487..2640
FT /tag= m
FT /number= 6
FT 2641..2910
FT /tag= n
FT /number= 6
FT 2911..2998
FT /tag= o
FT /number= 7
FT 2999..3284
FT /tag= p
FT /number= 7
FT 3285..3389
FT /tag= q
FT /number= 8
FT 3390..3470
FT /tag= r
FT /number= 8
FT 3471..3592
FT /tag= s
FT /number= 9
FT 3593..3891
FT /tag= t
FT /number= 9
FT 3892..4042
FT /tag= u
FT /number= 10
FT 4043..4285
FT /tag= v
FT /number= 10
FT 4286..4351
FT /tag= w
FT /number= 11
FT 4352..4647
FT /tag= x
FT 4648..4737
FT /tag= y
FT /number= 12
FT
FT WO200026390-A2.
FT 11-MAY-2000.
FT
FT 28-OCT-1999; 99WO-US25452.
FT 29-OCT-1998; 98US-0106239.
FT
```


QY	1068	AGGGGATCAATATNCCAGACTCTGGCCGTAGACATGCTGAACACCAAGGCAATTTCCACGC	1127
Db	257	AGGGGGCTACCAATATACAGACCCTTGGTGTGGCCCAAGCTGAGAGAGAAAGGCAATTTCCGCG	316
QY	1128	ATTACCAACAGTTATACCTGCACAGATGATGATGGTCCAGCAATT--GGTGCAGCAACTTT	1185
Db	317	ATTACCAACAGTNCCTGGTACTGTTGATTCATTTAGAGANNGTNAGNTACGCCAGCTTT	376
QY	1186	ACAACATCGTAGATGTTG---CATGAGGTCCATGATCTTACTCATTTGGCAATTTTGTGAAA	1242
Db	377	ACAAGCTTATTTGATGGGTATACGAAAGTTCCATGACATTTACCCACCTACCTTTTGTGTGAA	436
QY	1243	GAGAACGATGATGCTGAT--TAGATTTGGCCGTGAACGGCTGCTGATAGAGAGATGTCCTGGAC	1301
Db	437	GGGAACCTTATTCCTTATCTTAAGAGTTTCTGTCCACACTGCTGCTGGAGAGAAATCTTACAT	496
QY	1302	ATTGCTAGTATTTTTCAGGGGCTTAAGCTGTTGACGATCTGATCAGACAAATTACTTTTGAC	1361
Db	497	ATTGCTCAAAATCTTCGAGGCAAAACCTGTTGATGTTTGTGACACACAGTAAAGCTTGAC	556
QY	1362	CTTACTGCGGATCTAAGCAAGATGGTTG-CACTGCAAGAGTTATTGGAGCCCTAT	1415
Db	557	CTTACTGAGATCTTGACAAAGATGGTTGACACTACACAAAGATTATTGGAGCCATAT	611
RESULT 10			
ABA98822			
ID	ABA98822	standard; DNA: 724 BP.	
XX	ABA98822:		
AC			
XX			
DT	18-JUN-2002	(first entry)	
XX			
DE	Wheat ALS small subunit (clone wdk2c.pk015.a13) coding sequence.		
XX			
KW	ALS: Acetolactate synthase; acetoaldehyde synthase;		
KW	amino acid biosynthesis; plant; herbicide; ALS small subunit;		
KW	ALS holonzyme; crop protection chemical; expressed sequence tag;		
KW	EST; ss.		
XX			
OS	Triticum aestivum.		
XX			
XX	Key	Location/Qualifiers	
FT	CDS	2..403	
FT		/+tag= a	
FT		/partial	
FT		/product= "ALS small subunit"	
FT		/EC_number= "4.1.3.18"	
FT		/note= "no start codon present"	
XX			
PN	US2001044939-A1.		
XX			
PD	22-NOV-2001.		
XX			
XX	08-DEC-2000; 2000US-0732618.		
PF			
XX	04-JAN-2000; 2000US-174437P.		
PR			
XX			
PA	(ABEL/.) ABELL L M.		
PA	(FALC/) FALCO S C.		
PA	(FAMO/) FAMODU O O.		
XX			
PI	Abell LM, Falco SC, Famodu OO:		
XX			
DR	WPI: 2002-113395/15.		
DR	P-PSDB: ABB08418.		
XX			
PT	New small subunits of plant acetolactate synthase and nucleic acids		
PT	encoding them, useful in screening for novel crop protection chemicals		
PT	or potential herbicidal compounds based upon holonzyme inhibition		
PS	Claim 7; Page 40; 57pp: English.		

[illegible]

Query Match	Best Local Similarity	Score	DB	Length	Matches	Conservative	Mismatches	Indels	Gaps	
QY	460	AGCAGTCATCGAGCACTCCAGAGCTCGTAAAGTCTTAAAGTTGGAGATATCTCA	519	1	ACCAAGTCATTGAGCAGCTCAATAAGCTCGTAAAGCTTCAATGTTGAAAGATCTATCA	60	579	580	GGGCGAGATCATGTCGCTAGTTGACACATTCAGAGCAAGAGTTGTAGTATAGCGGAAC	639
Db	61	AAGAAGCTTCAGATTGAAAGAGAGCTGATGCTTATTAAGCTTAACGNNAGCACTGATCAC	120	121	GCCGAGAGCATGGTTTATGTCACATTTTCAGAGCAAAATNTGNTATANAAGTGA	180				

OY	640	ATGCTTTCAGTACGAGTA	CGTAGTCCGTGGAAAAATGATTGC	TGAAGAAGAAATT	699				
Db	181	AAACACTTACCANNAGAGN	AGCCTGGACATCCTGCCAAAATTGCTGC	AGTCCAGAGCATC	240				
OY	700	TGAAAAAGTTTCAGATC	GAGAGATTGTAAAGACAGGAAAGATP	AGCACTGAGAAGGAAA	759				
Db	241	TAAAGAAATTCGGCAT	CANAGAAATTTTGACAGACAGGAAAAAT	TGCTTGAACGACGNAAA	300				
OY	760	AGATGGGTGCACAAC	TGCTCATTTTGGCGATTTTTACGACAGCAT	CTTATCCAGATCTCAAG	819				
Db	301	AGATTGGNGCAAACAC	CCCCGTTTTCGGCGATTTCGCTGCTT	CTTATCCAGACCTTAAG	360				
OY	820	AGCAGAGCCCCTGT	GTGTCTTCGTGAAGTAGCAAAAAAGAG	AGCCATTGCTCCTCAAAAG	879				
Db	361	AGGCATTTCACAAAAA	AACCCGCTTACATCTGNAAAATTAAG	ACAGTGAATGCGAGTTT	420				
OY	880	AAACATC--	-AGCAGGGGGAGATGTTATCCCGTTG	AGCCAT	918				
Db	421	GACCATTCCAATGCT	GGGGGTGATGTNATCCNGNGAATCTT		462				
<hr/>									
RESULT 12									
ID	ABA98817	ABA98817 standard; DNA; 515 BP.							
XX	AC	ABA98817;							
DT	18-JUN-2002	(first entry)							
DE	Rice	ALS small subunit (clone cr10.pk084.a24) EST fragment.							
KM	AL5:	Acetolactate synthase: acetohydroxy synthase:							
KW	amino acid biosynthesis; plant;	herbicide; ALS small subunit;							
KM	ALS holoenzyme; crop protection	chemical; expressed sequence tag;							
KW	EST; ss.								
OS	Oryza sativa.								
FH	Key	Location/Qualifiers							
FT	CDS	1..471							
FT		/*tag= a							
FT		/product= "ALS small subunit"							
FT		/EC.number= "4.1.3.18"							
FT		/note= "no start codon present"							
XX	PN	US2001044939-A1.							
XX	PD	22-NOV-2001.							
XX	PE	08-DEC-2000; 2000US-0732618.							
XX	PR	04-JAN-2000; 2000US-174437P.							
XX	PA	(ABELL/) ABELL L M.							
XX	PA	(FALCO/) FALCO S C.							
XX	PI	(FAMO/) FAMODU O O.							
XX	PT	Abell LM, Falco SC, Famodu OO;							
XX	DR	WPI: 2002-11395/15.							
XX	P-	PSDB; ABB08413.							
XX	PT	New small subunits of plant acetolactate synthase and nucleic acids							
XX	PT	encoding them, useful in screening for novel crop protection chemicals							
XX	PT	or potential herbicidal compounds based upon holoenzyme inhibition -							
XX	PS	Claim 7; Page 33; 57pp; English.							
CC	The invention relates to new isolated polypeptides, which comprises an								
CC	acetolactate synthase (ALS) small subunit. Acetolactate synthase is the								
CC	first committed step in branched chain amino acid biosynthesis in plants								

CC and bacteria. The polypeptides are useful in aiding the discovery of new
 CC herbicides that inhibit plant acetoacetate synthase activity. The
 CC polypeptides are useful for preparing plant ALS holoenzyme, which are
 CC useful in screening for potential herbicidal compounds based upon
 CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
 CC for novel crop protection chemicals. Particularly, the polypeptides can
 CC be used as a target to facilitate design and/or identification of
 CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
 CC polypeptides could be appropriate for new herbicide discovery and
 CC design. The polynucleotides are also useful for producing the
 CC polypeptides and for screening for new herbicides. The current sequence
 CC represents the rice ALS small subunit (clone r10.pk084.a24) EST
 CC fragment.

CC Sequence 515 BP; 141 A; 108 C; 119 G; 145 T; 2 other;

Query Match 10.1%; Score 169.4; DB 24; Length 515;
 Best Local Similarity 65.2%; Pred. No. 6.2e-42;
 Matches 311; Conservative 0; Mismatches 158; Indels 8; Gaps 4;

OY 568 ATCCAGATCCAGGCGAGATGATGCTGGTGTGACACATTCAGAGCAAGATTGTAG 627
 DB 11 AACCCAGATCAGGCTCTGAGGTCATGTTTACTGATATTTTCCAGAGCAAGATTGTG 70
 OY 628 ATATACCGGAGACATGCTATGATGAGGTAATCAGATCCTGGAAAAATGATTGCTG 687
 DB 71 ATATTTCGGAGAACCCCTTACCAGAGTACCTGGAGATCCTGGCAAAATTTGTTGCTG 130
 OY 688 TAGAAAAGAAATTTGAAAAGTTTCACATCAGAGAGATTGTAAGACAGCAAGATAGCAC 747
 DB 131 TGCAAAAGAACCTCAGCAAAATTTGGATTAAGAAATTTGTAGAACGCGAAAAATTTGCTT 190
 OY 748 TGAGAGGAAAAAGATGGGTCACATGCTCCATTTTGGGATTTTCCAGACATCTCTATC 807
 DB 191 TGAGAGCTGAAAAAATTTGGAGCACTGCCCTTCTGGGATTTTCTGCTCTTTTACC 250
 OY 808 CAGATCTCAAGAGCAAGGCGCTGTAGTGTCTTTCGAA---GTACCAAAAAAGAGCCCA 864
 DB 251 CAGATCTCAAGAGCAAGGCGCTGTAGTGTCTTCTTCTTCTGTAATTAACAGCTCA 310
 OY 865 TTGTCCCTCAAAAAGAAACATC---AGCAGGGGAGATGTTTATCCCGTTGAGCCATTCTT 921
 DB 311 ATGGAAGTTTGTATCAACATCCATGCTGGGGCGATGCTATCTCTGTGGAACCTTATG 370
 OY 922 TTGACCCCAAGTACATGCTATCTGAGGCTCAGTGGGAGCTTCTCAGCAGAGATTA 981
 DB 371 AGGCTTATCATTGACCAAGATGCTGCTGAGGCGCTCTGATGATGAAGAT- 429
 OY 982 CGAGTGAGCTACGGTGCATCTCTATCTGTTGTTGTAATGATATTCAGAGAGTTTC 1038
 DB 430 CAGGTGAGCTTCATCATCACTACCAACA-TCCNTGTCAATGATTTGCCCTGGAGTTTC 485

RESULT 13
 ID ABA98818 standard; DNA; 484 BP.

XX ABA98818;

DT 18-JUN-2002 (first entry)

DE Rice ALS small subunit (clone r10n.pk117.a16) EST fragment.

XX ALS: Acetoacetate synthase; acetylhydroxy synthase;
 KW amino acid biosynthesis; plant; herbicide; ALS small subunit;
 KW ALS holoenzyme; crop protection chemical; expressed sequence tag;
 KW EST; ss.

OS Oryza sativa.

XX Key Location/Qualifiers
 FH CDS 11..444
 FT /*tag= a

FT /partial
 FT /product= "ALS small subunit"
 FT /EC number= "4.1.3.18"
 FT /transl_except= (pos:350..352, aa:Xaa)
 FT /transl_except= (pos:419..421, aa:Xaa)
 FT /transl_except= (pos:422..424, aa:Xaa)
 FT /transl_except= (pos:434..436, aa:Xaa)
 FT /transl_except= (pos:440..442, aa:Xaa)
 FT /note= "Xaa is unknown; no start codon present"

PN US2001044939-A1.

XX 22-NOV-2001.

PF 08-DEC-2000; 2000US-0732618.

PR 04-JAN-2000; 2000US-174437P.

PA (ABELL) ABELL, L M.

PA (FALC/) FALCO S C.

PA (FAMO/) FAMODU O O.

PI Abell LM, Falco SC, Famodu OO;

DR WPI: 2002-113395/15.

XX P-PSDB; ABB08414.

PT New small subunits of plant acetoacetate synthase and nucleic acids
 PT encoding them, useful in screening for novel crop protection chemicals
 or potential herbicidal compounds based upon holoenzyme inhibition

PS Claim 7; Page 34; 57pp; English.

CC The invention relates to new isolated polypeptides, which comprises an
 CC acetoacetate synthase (ALS) small subunit. Acetoacetate synthase is the
 CC first committed step in branched chain amino acid biosynthesis in plants
 CC and bacteria. The polypeptides are useful in aiding the discovery of new
 CC herbicides that inhibit plant acetoacetate synthase activity. The
 CC polypeptides are useful for preparing plant ALS holoenzyme, which are
 CC useful in screening for potential herbicidal compounds based upon
 CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
 CC for novel crop protection chemicals. Particularly, the polypeptides can
 CC be used as a target to facilitate design and/or identification of
 CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
 CC polypeptides could be appropriate for new herbicide discovery and
 CC design. The polynucleotides are also useful for producing the
 CC polypeptides and for screening for new herbicides. The current sequence
 CC represents the rice ALS small subunit (clone r10n.pk117.a16) EST
 CC fragment.

SO Sequence 484 BP; 147 A; 103 C; 115 G; 112 T; 7 other;

Query Match 9.6%; Score 160.8; DB 24; Length 484;
 Best Local Similarity 66.3%; Pred. No. 2.9e-39;
 Matches 246; Conservative 0; Mismatches 123; Indels 2; Gaps 1;

OY 394 CTTTGTGCTGAGACAGAGAAAGGCTCTATTACCACTAGTGTCTGAGCACTGAAGGG 453
 DB 12 CCGTGGGGCTCAACAAGAGCAAGGCGCATGTTTCACATGCTGCTCCGAGCGAGAGGG 71
 OY 454 TACTTACGAGGTCATCAGACCACTCAGAGGCTGTTAATGTTCTTAAAGGTTGAAGATA 513
 DB 72 TGCTCAACCAAGTCATCAGACACAGCTTGTCAACGCTTGAATGATGGAAGATC 131
 OY 514 TCTCAAGTGAGCCGCAAGTGGAGCTGATGCTTGTAAAGTGATGCACATCCAG 573
 DB 132 TATCTAAGGAGCCACAGGTTGAAAGAGAGCTGATGCTTATTAATAATGTTGAACCA 191
 OY 574 AATCCAGGGCAGAGATCATTGCTGCTGACACATTCAGAGCAAGAGTGTGATATAG 633
 DB 192 ATCAGGCTCTGAGGTCATGAGTTTGTAGTGAATATTTTCCGAGCAAGAGTGTGATATAT 251
 OY 634 CGGACATGCTATGATCAGAGTAAGTGG--AGATCTCGAATAATGATGCTGTAGA 691

PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.

PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 03-AUG-1999; 990S-0146389.
PR 04-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 05-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155139.
PR 24-SEP-1999; 990S-0155486.
PR 28-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.

```
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.
```

```
Query Match      8.8%; Score 147.6; DB 21: Length 420;
Best Local Similarity 69.5%; Pred. No. 3.3e-35;
Matches 198; Conservative 1; Mismatches 86; Indels 0; Gaps 0;

QY 1234 TTTCGAAAGAGACTGATGCTGATTAGATTCGCCGTAACGCTGCTGCTAGAGAGATG 1293
    ||| : || ||||| ||| ||| ||| ||| |||||
DB   8 TTGCGGGMGAMNACTGATGCTCATCAAGTACGCTGCTGATCTCTGCAAGGAGATG 67

QY 1294 TCCTGGACATTTGCTAGTATTTTCAGGGCTPAAGCTGTGACGTATCTGATCACAATTA 1353
    | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   68 TTTTAGATATGCTCAAGTTTCCGTGCTPAAGCCATGTGATGTCCTGATCATACTATCA 127

QY 1354 CTTTGCAGCTTACTGGGAGATCTAGACAGATGCTGCACCTGCAAGGTTATTGAGCCCT 1413
    | | || ||||| ||||| || ||| ||||| ||||| || ||| |||
DB   128 CCCTTGAGGTTAGAGGGATCTCAGAAAATGTCGTGACTACAAAGCAGTTAGAGCCT 187

QY 1414 ATGCTATATGTGAGTTCAGAGAACCGGTCGTGTCATTTGGCTCGTAATCGGAGTGG 1473
    |||| || ||||| || ||| || ||||| || ||||| ||||| |||||
DB   188 ATGGAATCTGCGAGTGCCTCGGACAGAGAGTGGCATTGGTAAGAGAACTGGAGTGG 247

QY 1474 ACTCAAGTACCTTCGGGATACTCCTTCTTTTAAAGGCTAA 1518
    | ||| || ||||| ||||| ||||| ||||| |||||
DB   248 ATTCGACTTATCTTCGGGATACTCTCTCTTGTAGTTCTAAA 292
```

Search completed: March 9, 2003, 19:39:42
Job time : 444 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 09:00:40 ; Search time 40 Seconds

(without alignments)
1635.651 Million cell updates/sec

Title: US-09-997-900-2

Perfect score: 2444

Sequence: 1 MAAISVSSPSIRCLRSACS.....ARESGVDKYLRGYSFLLTG 491

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_101002:*

- 1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2444	100.0	491	21	AA196203
2	2358.5	96.5	484	23	ABB91968
3	1692	69.2	476	23	ABB08416
4	1518	62.1	477	23	ABB93502
5	1493.5	61.1	483	23	ABB08412
6	1476	60.4	449	19	AAW79140
7	1395	57.1	365	23	ABB08415
8	1302	53.3	350	23	ABB08410
9	522	21.4	203	23	ABB08408
10	499	20.4	133	23	ABB08418

11	470.5	19.3	156	23	ABB08409
12	436	17.8	156	23	ABB08413
13	372.5	15.2	144	23	ABB08414
14	336	13.7	94	21	AA16178
15	332	13.6	86	21	AA16179
16	329.5	13.5	135	23	ABB08417
17	329.5	13.5	174	22	AA01249
18	310.5	12.7	803	22	AA831992
19	307	12.6	172	22	AA933204
20	305	12.5	166	22	AA037779
21	305	12.5	166	22	AA038002
22	305	12.5	186	17	AAW06555
23	302.5	12.4	158	19	AAW05930
24	301	12.3	172	22	AA879360
25	277.5	11.4	163	22	AA667740
26	277	11.3	163	22	AA036451
27	276.5	11.3	163	22	AA034430
28	271.5	11.1	173	22	AA038425
29	269.5	11.0	173	22	AA036149
30	268.5	11.0	163	22	AA035663
31	267.5	10.9	163	22	AA035456
32	263.5	10.8	163	23	ABB48167
33	262.5	10.7	158	23	ABB54557
34	260.5	10.7	158	15	AA854221
35	138	5.6	43	23	ABB08411
36	137.5	5.6	104	22	AA053173
37	135	5.5	828	22	AA931284
38	134.5	5.5	97	23	ABP07823
39	121	5.0	1138	14	AA837213
40	115	4.7	1583	22	AA053954
41	112	4.6	1138	15	AA846225
42	111	4.5	495	22	AA034722
43	110.5	4.5	737	21	AA09614
44	110	4.5	507	22	AA86433
45	107	4.4	500	22	AA038171

ALIGNMENTS

RESULT 1	AA196203	standard; Protein; 491 AA.
ID	AA196203	
AC	AA196203	
DT	11-AUG-2000	(first entry)
DE	Arabidopsis acetohydroxy-acid synthase (AHAS) small subunit.	
XX	Herbicide-resistance; acetohydroxy-acid synthase; AHAS;	
XX	acetolactate synthase; imidazolone; sulfonylurea;	
XX	triazolopyrimidine sulfonamide; sulfamoylurea; enzyme;	
XX	pyrimidyl-oxy-benzolic acid; sulfonycarboxamide; transgenic plant;	
XX	branched-chain amino acid synthesis.	
OS	Arabidopsis sp.	
XX	WO200026390-A2.	
XX	PD	11-MAY-2000.
XX	PF	28-OCT-1999; 99WO-US25452.
XX	PR	29-OCT-1998; 98US-0106239.
XX	PA	(AMCY) AMERICAN CYANAMID CO.
XX	PI	kakefuda G, Costello C, Sun M, Hu W;
XX	DR	WPI; 2000-365633/31.
XX	XX	N-PSDB; AAA27424, AAA27425.

Corn ALS small sub
Rice ALS small sub
Rice ALS small sub
Arabidopsis thalia
Arabidopsis thalia
Wheat ALS small su
B. subtilis acetoh
Amino acid sequenc
C glutamicum prote
Streptococcus pneu
Streptococcus pneu
R. capsulatae acet
S. pneumoniae deri
Corynebacterium q1
Small subunit of i
Pseudomonas aerugi
E. coli cellular p
Salmonella typhi c
Klebsiella pneumon
Haemophilus influe
Klebsiella pneumon
Listeria monocytog
Lactococcus lactis
L.lactis alpha-ace
Corn ALS small sub
Propionibacterium
C glutamicum prote
Human ORFX protein
B.t. toxin HD311.
Propionibacterium
Bacillus thuringie
E. coli cellular p
Streptococcus pneu
putative P. abyssi
Salmonella typhi c

PT New polynucleotide encoding eukaryotic acetoaldehyde-synthetase
 PT small subunit protein for producing transgenic herbicide resistant
 PT plants and identifying mutations affecting enzymatic activity of the
 PT synthetase

PS Disclosure: Page 50-52; 57pp; English.

CC The present sequence is the Arabidopsis acetoaldehyde-synthetase (AHAS)
 CC small subunit protein. AHAS is also known as acetolactate synthase. This
 CC enzyme is needed for branched-chain amino acid synthesis and so is
 CC essential for life. Inhibition of this enzyme would lead to plant death
 CC and therefore inhibitors would be potential herbicides. Certain
 CC herbicides are known to inhibit AHAS: imidazolinones, sulfonylureas,
 CC triazopyrimidine sulfonamides, pyrimidyl-oxy-benzoic acids,
 CC sulfamoylureas and sulfonycarboxamides. The present sequence may be used
 CC to identify mutant AHAS which are resistant to these herbicides and
 CC may be used to create herbicide resistant transgenic plants e.g. dicot
 CC and monocot crop plants.

CC Sequence 491 AA:

Query Match 100.0%; Score 2444; DB 21; Length 491;
 Best Local Similarity 100.0%; Pred. No. 2e-220;
 Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAISVSSPSIRCLRSACSDSPALVSSTRVSFPAKISYLSGISHRDEMGKMEGCV 60
 DB 1 MAISVSSPSIRCLRSACSDSPALVSSTRVSFPAKISYLSGISHRDEMGKMEGCV 60
 QY 61 RSVGKISDASFSSEASAPPKSKVRKHTISVFVGDESGMINRIAGFARGYNIESLAVG 120
 DB 61 RSVGKISDASFSSEASAPPKSKVRKHTISVFVGDESGMINRIAGFARGYNIESLAVG 120
 QY 121 LNRDALFTIVVCGTGRVLOQYIEOLQKLVNLYLKVEDISSEPOVERELMLVYNNAHPESR 180
 DB 121 LNRDALFTIVVCGTGRVLOQYIEOLQKLVNLYLKVEDISSEPOVERELMLVYNNAHPESR 180
 QY 181 AEIMMLVDFRFRVVDIAEHALTIEVTGDPGKMI AVERNLLKFFQIREIVRTGKIALRREK 240
 DB 181 AEIMMLVDFRFRVVDIAEHALTIEVTGDPGKMI AVERNLLKFFQIREIVRTGKIALRREK 240
 QY 241 MCATAPFMEFSAASYPDLKEQAPVSVLRSSKGAIVPOKETSAGGCVVVEPFDPKVR 300
 DB 241 MCATAPFMEFSAASYPDLKEQAPVSVLRSSKGAIVPOKETSAGGCVVVEPFDPKVR 300
 QY 301 IIDAHHGLTDEDTGSLRSHLTLVNDIPGVNLITVGVFARGYNIOQLAVGHAETGCI 360
 DB 301 IIDAHHGLTDEDTGSLRSHLTLVNDIPGVNLITVGVFARGYNIOQLAVGHAETGCI 360
 QY 361 SRITTVIPATDES SVSKLYOOLYKLVNDVHEVHDLTHLPSEERELMLIKIAVNAARADVLD 420
 DB 361 SRITTVIPATDES SVSKLYOOLYKLVNDVHEVHDLTHLPSEERELMLIKIAVNAARADVLD 420
 QY 421 IASIFRAKAVDVSDHTITTLQTLGDLKMYALORLLEPGICVARTGRVALARESGVDSK 480
 DB 421 IASIFRAKAVDVSDHTITTLQTLGDLKMYALORLLEPGICVARTGRVALARESGVDSK 480
 QY 481 YLRGYSFLLTG 491
 DB 481 YLRGYSFLLTG 491

RESULT 2
 ABB91968

ID ABB91968 standard: Protein; 484 AA.

AC ABB91968;

DT 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 1179.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.
 OS
 XX
 PN WO200210210-A2.
 XX

PD 07-FEB-2002.

PF 28-AUG-2001: 2001WO-EP09892.

PR 28-AUG-2001: 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tiejfen K, Weidler M;

DR WPI: 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -

PS Claim 5; SEQ ID NO 1179; 261pp + Sequence Listing; English.

CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

CC Sequence 484 AA:

Query Match 96.5%; Score 2358.5; DB 23; Length 484;
 Best Local Similarity 97.2%; Pred. No. 2.1e-212;
 Matches 480; Conservative 0; Mismatches 1; Indels 13; Gaps 2;

QY 1 MAISVSSPSIRCLRSACSDSPALVSSTRVSFPAKISYLSGISHRDEMGKMEGCV 60
 DB 1 MAISVSSPSIRCLRSACSDSPALVSSTRVSFPAKISYLSGISHRDEMGKMEGCV 60
 QY 61 RSVGKISDASFSSEASAPPKSKVRKHTISVFVGDESGMINRIAGFARGYNIESLAVG 120
 DB 61 RSVGKISDASFSSEASAPPKSKVRKHTISVFVGDESGMINRIAGFARGYNIESLAVG 120
 QY 121 LNRDALFTIVVCGTGRVLOQYIEOLQKLVNLYLKVEDISSEPOVERELMLVYNNAHPESR 180
 DB 121 LNRDALFTIVVCGTGRVLOQYIEOLQKLVNLYLKVEDISSEPOVERELMLVYNNAHPESR 180
 QY 181 AEIMMLVDFRFRVVDIAEHALTIEVTGDPGKMI AVERNLLKFFQIREIVRTGK ---IALR 237
 DB 181 AEIMMLVDFRFRVVDIAEHALTIEVTGDPGKMI AVERNLLKFFQIREIVRTGKVVYIALR 240
 QY 238 REKMGATAPFMEFSAASYPDLKEQAPVSVLRSSKGAIVPOKETSAGGCVVVEPFDPK 297
 DB 241 REKMGATAPFMEFSAASYPDLKEQAPVSVLRSSKGAIVPOKETSAGGCVVVEPFDPK 300
 QY 298 VARIIDAHHGLTDEDTGSLRSHLTLVNDIPGVNLITVGVFARGYNIOQLAVGHAET 357
 DB 301 VARIIDAHHGLTDEDTGSLRSHLTLVNDIPGVNLITVG-----SLAVGHAET 350
 QY 358 KGISRTTVIPATDES SVSKLYOOLYKLVNDVHEVHDLTHLPSEERELMLIKIAVNAARADVLD 417
 DB 351 KGISRTTVIPATDES SVSKLYOOLYKLVNDVHEVHDLTHLPSEERELMLIKIAVNAARADVLD 410
 QY 418 VLDIASIFRAKAVDVSDHTITTLQTLGDLKMYALORLLEPGICVARTGRVALARESGV 477
 DB 411 VLDIASIFRAKAVDVSDHTITTLQTLGDLKMYALORLLEPGICVARTGRVALARESGV 470
 QY 478 DSKYLRGYSFLLTG 491

```
Db      471 DSKYLGYSFPLTG 484
|||||
RESULT 3
ABB08416
ID      ABB08416 standard; Protein; 476 AA.
XX
XX      ABB08416;
XX
XX      18-JUN-2002 (first entry)
XX
DE      Soybean ALS small subunit (clone sdc2c.pk001.b10).
XX
XX      ALS; Acetolactate synthase; acetohydroxy synthase;
XX      amino acid biosynthesis; plant; herbicide; ALS small subunit;
XX      ALS holoenzyme; crop protection chemical; enzyme.
XX
XX      Glycine max.
XX
XX      US2001044939-A1.
XX
XX      22-NOV-2001.
XX
XX      08-DEC-2000; 2000US-0732618.
XX
XX      04-JAN-2000; 2000US-174437P.
XX
XX      (ABEL/) ABELL L M.
XX      (FALC/) FALCO S C.
XX      (FAMO/) FAMODU O O.
XX
XX      Abell LM, Falco SC, Famodu OO;
XX
XX      WPI: 2002-113395/15.
XX      N-PSDB: ABA98820.
XX
XX      New small subunits of plant acetolactate synthase and nucleic acids
XX      encoding them, useful in screening for novel crop protection chemicals
XX      or potential herbicidal compounds based upon holoenzyme inhibition -
XX
XX      Claim 19; Fig 3; 57pp; English.
XX
XX      The invention relates to new isolated polypeptides, which comprises an
XX      acetolactate synthase (ALS) small subunit. Acetolactate synthase is the
XX      first committed step in branched chain amino acid biosynthesis in plants
XX      and bacteria. The polypeptides are useful in aiding the discovery of new
XX      herbicides that inhibit plant acetolactate synthase activity. The
XX      polypeptides are useful for preparing plant ALS holoenzyme, which are
XX      useful in screening for potential herbicidal compounds based upon
XX      holoenzyme inhibition. The ALS holoenzyme is also useful in screening
XX      for novel crop protection chemicals. Particularly, the polypeptides can
XX      be used as a target to facilitate design and/or identification of
XX      inhibitors of those enzymes that may be useful as herbicides. Thus, the
XX      polypeptides could be appropriate for new herbicide discovery and
XX      design. The polynucleotides are also useful for producing the
XX      polypeptides and for screening for new herbicides. The current sequence
XX      represents the soybean ALS small subunit (clone sdc2c.pk001.b10).
XX
SQ      Sequence 476 AA:
Query Match      69.2%; Score 1692; DB 23; Length 476;
Best Local Similarity 71.5%; Pred. No. 7,1e-150;
Matches 337; Conservative 57; Mismatches 69; Indels 8; Gaps 2;
```

```
QY      139 LQVIEQLQKLVNVLKVEDISSEPOVERELMLVKYNAHDESPRAEIMWLVDFFARVVDIA 198
      1:|||||
Db      126 LQVMEQLQKLVNVLKVEDLSREPQVERELMLIKVHADPKHHAELKWLVDIFRAKIVDIS 185
QY      199 EHALLTEVTGDPGKMAVRNKKFOIREIVTCKIALRREKMGATAPWRSSAASPDL 258
      1:|||||
Db      186 EHSVLTETVGDGKMAAAYORNRKFKGIKEIARTGKIALRREKMGASAPWRSSAASPDL 245
QY      259 KEQAPSVLRSKKGAIYQKETSAGGDVYPVEPFDPVHRLDLAHWGLLDEPDSGLR 318
      2:|||||
Db      246 EGRTPYNALVGAKNNKPVAKLDTPVGGDVYPIEPSDGSVNOVLDAHMGVLNDEDSGLR 305
QY      319 SHTLSLVNDIEGVNIYGVFARRGYNQISLAVGHAETKGISRTTYPATDESYSKLV 378
      1:|||||
Db      306 SHTLSLVNDAPGVNIYGVFARRGYNQISLAVGHAETKGISRTTYPATDESYSKLV 365
QY      379 QOLYKLVDEVHVDLTHLPFSERELMLIKIYVNAARVDLIASIFRAKAVDSDHTTT 438
      1:|||||
Db      366 QOLYKLVDELHEVVDITHLPFAERELMLIKIYVNAARVDLIASIFRAKAVDSDHTTT 425
QY      439 LQITGDLDKMVALQRLPEYIGICEVARTGRVALARESGVDSKYLGYSPFL 489
      1:|||||
Db      426 LELTGDLDKMVALQRLPEYIGICEVARTGRVALARESGVDSKYLKGYSTPL 476
```

```
RESULT 4
ABB93502
ID      ABB93502 standard; Protein; 477 AA.
XX
XX      ABB93502;
XX
XX      31-MAY-2002 (first entry)
XX
XX      Herbicidally active polypeptide SEQ ID NO 2713.
XX
XX      Herbicidal; plant; agriculture; herbicide.
XX
XX      Arabidopsis thaliana.
XX
XX      WO200210210-A2.
XX
XX      07-FEB-2002.
XX
XX      28-AUG-2001; 2001WO-EP09892.
XX
XX      28-AUG-2001; 2001WO-EP09892.
XX
XX      (FARB ) BAYER AG.
XX
XX      Tietjen K, Weidner M;
XX
XX      WPI: 2002-269010/31.
XX
XX      Identifying plant target proteins for herbicidally active compounds,
XX      comprising aligning and comparing nucleic acid or amino acid sequences
XX      from plant with nucleic acid or amino acid sequences from non-plant
XX      organisms -
XX
XX      Claim 5; SEQ ID NO 2713; 261pp + Sequence Listing; English.
XX
XX      The invention relates to identifying target proteins
XX      (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX      aligning and comparing nucleic acid or amino acid sequences from plant
XX      with nucleic acid or amino acid sequences from non-plant organisms using
XX      suitable search parameters, where plant sequences having an E-value
XX      greater by a factor of 3 than the E-value of most similar non-plant
XX      sequences are selected. The polypeptides or nucleic acids encoding them
XX      are useful for identifying modulators. The identified modulators are
XX      useful as herbicides.
XX
SQ      Sequence 477 AA:
Query Match      62.1%; Score 1518; DB 23; Length 477;
```

Best Local Similarity 69.8%; Pred. No. 1.5e-133;
Matches 296; Conservative 59; Mismatches 65; Indels 4; Gaps 2

[illegible]

RESULT 5	ABBO8412	ABBO8412 standard; protein: 483 AA.
XX	ID	ABBO8412
XX	AC	ABBO8412:
XX	DT	18-JUN-2002 (first entry)
XX	DE	Corn ALS small subunit (clone p0094.csst172a).
XX	KW	ALS: Acetolactate synthase; EC 4.1.3.18; acetohydroxy synthase;
KM	amino acid biosynthesis; plant; herbicide; ALS small subunit;	
KW	ALS holoenzyme; crop protection chemical; enzyme.	
XX	Zea mays.	
OS	US2001044939-A1.	
XX	PN	22-NOV-2001.
XX	PD	08-DEC-2000; 2000US-0732618.
XX	PE	04-JAN-2000; 2000US-174437P.
XX	PR	(ABELL) ABELL L M.
XX	PA	(FALC/) FALCO S C.
XX	PA	(FAMO/) FAMODU O O.
XX	PI	Abell LM, Falco SC, Famodu OO:
XX	WPI:	2002-113395/15.
XX	DR	N-PSDB: ABA98816.
XX	PT	New small subunits of plant acetolactate synthase and nucleic acids
XX	PT	encoding them, useful in screening for novel crop protection chemicals
XX	or potential herbicidal compounds based upon holoenzyme inhibition	

PS Claim 19; Fig 3; 57pp; English.

The invention relates to new isolated polypeptides, which comprises an acetoacetate synthase (ALS) small subunit. Acetoacetate synthase is the first committed step in branched chain amino acid biosynthesis in plant and bacteria. The polypeptides are useful in aiding the discovery of new herbicides that inhibit plant acetoacetate synthase activity. The polypeptides are useful for preparing plant ALS holoenzyme, which are useful in screening for potential herbicidal compounds based upon holoenzyme inhibition. The ALS holoenzyme is also useful in screening for novel crop protection chemicals. Particularly, the polypeptides can be used as a target to facilitate design and/or identification of inhibitors of those enzymes that may be useful as herbicides. Thus, the polypeptides could be appropriate for new herbicide discovery and design. The polynucleotides are also useful for producing the polypeptides and for screening for new herbicides. The current sequence represents the corn ALS small subunit (clone p0094.cssl172a).

SQ Sequence 483 AA;

Query Match	61.18;	Score 1493.5;	DB 23;	Length 483;
Best Local Similarity	63.58;	Pred. No. 3.1e-131;		
Matches 308;	Conservative 61;	Mismatches 83;	Indels 33;	Gaps 7

Qy	13	RCMSACSDSSPALVSSNRVFEFPKXISYLSGSISSHHROGEMKRRHEGVRISVDGKISPA5F	72
Db	24	RCGAGSRVDSRPL---TPAVGFTA-----GPR---AR5V---AVTAA55	58
Qy	73	SEAS-SATP-----KSKVRKHTISVPEGDESGMINRIAGVFARGYNI5ESLAVGLNRDK	125
Db	59	SPATGGVTPVPPRNSR5VKKRHTLSVEFGDESGMINRIAGVFARGYNI5ESLAVGLNRDK	118
Qy	126	ALFTTVGCTEVLVOOVTEOLKLVNVLKVEDISSEPOVEKELLVYNAHPSERAEIMV	189
Db	119	ALFTTVASGCTILNQVVEOLKLVNVIKVDL5EMEPQVEKELMLIKVNAEREKLP5IMG	178
Qy	186	LVDFEFAFVNDIAEHALTIEVTSQDPGKMI5AVERNLLKFOIREIVRTGKIALRREKMGATA	245
Db	179	LVRIEFAFVNDLSVYTLTIEVTSQDPGKMI5AOKL5YKGI5REIARTGKIALRREKMGETA	238
Qy	246	PEWFR5AS5YDLDEQAFVSVLS5SKKGAIVPQKETSAGGVYVPEPFD-PKVHRI1LDA	304
Db	239	PEWFR5AS5YDLEVALP5NFQOVTGKAIQNP5ESSGGDVYVE5EST5S5TQLDA	296
Qy	305	HWGILLTDEDT5GLNSHTLSLLVNDIPGVLLVITGVFARGYNIQ5SLAVGAETKGISRIT	364
Db	299	HWCVTOTDDP5GFC5HTLSILVNVDPGVLVTVTSF5SRGYNIO5SLAVGPAEK5GTSRIT	356
Qy	365	TVI5PTD5SV5KIVQOVLKVLVDV5EHVHDTLPS5E5EMLIKI5AVNAAR5DVLDA5I	424
Db	359	TVI5PTD55IKVLHQLKTLIDV5EVODFTLPL5AAR5LMI5IK5AANA5TAR5DVLDA5I	418
Qy	425	FRAKAVND5DHTITLQTLGDLDDK5VVALQ5RL5EP5GICEVARTG5VAL5ARE5GVSKYLRG	484
Db	419	FEAQGVNDISDHTITLTLGLDIDR5VVLQ5KML5EQ5ICEVARTG5IALLR5GVDSKYLRG	478
Qy	485	YSFLL 489	
Db	479	FSLP. 483	

RESULT 6
AAW79140
ID AAW79140 standard; Protein; 449 AA.
XX
XX AAW79140;
AC
XX
DT 18-NOV-1998 (first entry)
XX
XX Plant acetoacetyl synthase (ALS) small subunit protein.
DE
XX
XX ALS: small subunit; acetoacetyl synthase; plant; ssuALS; lsuALS;
KW Nicotiana glauca; Nicotiana glauca; plasmid vector; hecB; holenzym;

KW large subunit.
 XX
 OS ~~Nicotiana glauca~~ *Nicotiana glauca*.
 XX
 PN MO9837206-A1.
 XX
 PD 27-AUG-1998.
 XX
 PE 23-FEB-1998; 98MO-US03506.
 XX
 PR 24-FEB-1997; 97US-0039148.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 XX
 PI Abell LM, Hershey HP;
 XX
 DR WPI: 1998-467568/40.
 XX
 DR N-PSDB: AAV35857.
 XX
 PR New nucleic acid encoding a plant aceto:lactate synthase small
 PT sub:unit - that combines with the large sub:unit to give a
 PT holoenzyme having higher activity than the large sub:unit alone
 XX
 PS Claim 1; Pages 30-32; 47pp: English.
 XX
 CC This represents a small subunit of a plant acetylactate synthase
 CC (ssuALS). The encoding cDNA is contained in the plasmid pSSU.NPI. The
 CC plasmid vector comprising the ssuALS nucleotide operably linked to a
 CC regulatory sequence can be used to transform host cells for the
 CC recombinant production of the protein which is used for evaluating a
 CC compound for acetylactate synthase inhibition and so for selecting
 CC potential herbicides. Evaluation of a compound for acetylactate synthase
 CC (ALS) inhibition comprises expression and purification of plant ssuALS
 CC from the transformed host, mixing this ssuALS with the large subunit of
 CC ALS to form a holoenzyme which is then treated with a test compound.
 CC Treated and untreated holoenzyme activity are compared to select
 CC compounds with potential for herbicidal activity. Previously, large
 CC subunits of plant ALS (lsuALS) have been isolated, but the existence of
 CC ssuALS had not been verified. Mixing the two subunits results in a
 CC holoenzyme that has a 4-15 fold increase in specific activity over
 CC lsuALS alone.
 XX
 SQ Sequence 449 AA:
 Query Match 60.4%; Score 1476; DB 19: Length 449;
 Best Local Similarity 79.1%; Pred. No. 1,2e-125;
 Matches 291; Conservative 95; Mismatches 42; Indels 0; Gaps 0;
 QY 99 MINRIAGVFARRGYNIESLAVGLNRDKALFTIVCGSTERVLQOVILOLQKLVNVLKVEDI 158
 DB 37 MINRIAGVFARRGYNIESLAVGLNRDKALFTIVCGSTERVLQOVILOLQKLVNVLKVEDL 96
 QY 159 SESEPOVERELMLVKVNAHESRAELMTLVTRARVVDIAEHALTEVTGDPGKMAVER 218
 DB 97 SESEPOVERELMLIKISADPKYAEVWMLVDFRAKVIDISDLSLETVEYGDGKMAVOR 156
 QY 219 NKKFQIRIVTGTALREKMGATAPWRFSASYPDLKQAPVSVRSKKGATVPQ 278
 DB 157 NKKFQIRIVTGTALREKMGATAPWRFSASYPDLKQAPVSVRSKKGATVPQ 216
 QY 279 KETSAGGDVYVPEFDPKVRHRLDAHMGILTDEDTSGLRSHTLSLVNDIGVLTIVTG 338
 DB 217 SMSMAGDGVYVPEFDPKVRHRLDAHMGILTDEDTSGLRSHTLSLVNDIGVLTIVTG 276
 QY 339 VVARGYNQOSLAVGHAETKGISRTITVPATDESYSKLVQOLYKLVNVEHVDLTHLDF 398
 DB 277 VVARGYNQOSLAVGHAETKGISRTITVPATDESYSKLVQOLYKLVNVEHVDLTHLDF 336
 QY 399 SERELMLIKIIVANAARVDLIASIFRAKAVDSOHTTLTGLTGDLMVALORLLEPY 458
 DB 337 SERELMLIKIIVANAARVDLIASIFRAKAVDSOHTTLTGLTGDLMVALORLLEPY 396
 QY 459 GICEVART 466

DB 397 GICEVART 404
 |||||||
 RESULT 7
 ABB08415
 ID ABB08415 standard; protein; 365 AA.
 XX
 AC ABB08415;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Rice ALS small subunit (clone r10npk117.a16).
 XX
 KW ALS: Acetylactate synthase; EC 4.1.3.18; acetylhydroxy synthase;
 KW amino acid biosynthesis; plant; herbicide; ALS small subunit;
 KW ALS holoenzyme; crop protection chemical; enzyme.
 XX
 OS Oryza sativa.
 XX
 PN US2001044939-A1.
 XX
 PD 22-NOV-2001.
 XX
 PF 08-DEC-2000; 2000US-0732618.
 XX
 PR 04-JAN-2000; 2000US-174437P.
 XX
 PA (ABELL) ABELL L M.
 PA (FALCO) FALCO S C.
 PA (FAMO) FAMODU O O.
 XX
 PI Abell LM, Falco SC, Famodu OO;
 XX
 DR WPI: 2002-113395/15.
 XX
 DR N-PSDB: ABA98819.
 XX
 PT New small subunits of plant acetylactate synthase and nucleic acids
 PT encoding them, useful in screening for novel crop protection chemicals
 PT or potential herbicidal compounds based upon holoenzyme inhibition
 XX
 PS Claim 19; Page 36; 57pp: English.
 XX
 CC The invention relates to new isolated polypeptides, which comprises an
 CC acetylactate synthase (ALS) small subunit. Acetylactate synthase is the
 CC first committed step in branched chain amino acid biosynthesis in plants
 CC and bacteria. The polypeptides are useful in aiding the discovery of new
 CC herbicides that inhibit plant acetylactate synthase activity. The
 CC polypeptides are useful for preparing plant ALS holoenzyme, which are
 CC useful in screening for potential herbicidal compounds based upon
 CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
 CC for novel crop protection chemicals. Particularly, the polypeptides can
 CC be used as a target to facilitate design and/or identification of
 CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
 CC polypeptides could be appropriate for new herbicide discovery and
 CC design. The polynucleotides are also useful for producing the
 CC polypeptides and for screening for new herbicides. The current sequence
 CC represents the rice ALS small subunit (clone r10n.pk117.a16).
 XX
 SQ Sequence 365 AA:
 Query Match 57.1%; Score 1395; DB 23: Length 365;
 Best Local Similarity 75.6%; Pred. No. 3.6e-122;
 Matches 276; Conservative 38; Mismatches 49; Indels 2; Gaps 1;
 QY 127 LFTIIVCGTERVLOQVIELOLQKLVNVLKVEDISEPOVERELMLVKVNAHESRAELIML 186
 DB 1 MFTIIVSGDRVNLQVIELOLQKLVNVLKVEDISEPOVERELMLIKINVEPQREPVWL 60
 QY 187 VDFRARRVVDIAEHALTEVTGDPGKMAVERNLKFKQIRIVTGTALREKMGATAP 246
 DB 61 VDFRARRVVDIAEHALTEVTGDPGKMAVERNLKFKQIRIVTGTALREKMGATAP 120

XX Abell LM, Falco SC, Famodu OO;
PI WPI: 2002-113395/15.
XX New small subunits of plant acetolactate synthase and nucleic acids
PT encoding them, useful in screening for novel crop protection chemicals
XX or potential herbicidal compounds based upon holoenzyme inhibition
PS Disclosure; Page 25-26; 57pp; English.
XX The invention relates to new isolated polypeptides, which comprises an
CC acetolactate synthase (ALS) small subunit. Acetolactate synthase is the
CC first committed step in branched chain amino acid biosynthesis in plants
CC and bacteria. The polypeptides are useful in aiding the discovery of new
CC herbicides that inhibit plant acetolactate synthase activity. The
CC polypeptides are useful for preparing plant ALS holoenzyme, which are
CC useful in screening for potential herbicidal compounds based upon
CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
CC for novel crop protection chemicals. Particularly, the polypeptides can
CC be used as a target to facilitate design and/or identification of
CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
CC polypeptides could be appropriate for new herbicide discovery and
CC design. The polynucleotides are also useful for producing the
CC polypeptides and for screening for new herbicides. The current sequence
CC represents the corn ALS small subunit (clone m15.12.b12.sk20).
XX Sequence 203 AA;
SQ
Query Match 21.4%; Score 522; DB 23; Length 203;
Best Local Similarity 60.1%; Pred. No. 1.3e-40;
Matches 107; Conservative 27; Mismatches 40; Indels 4; Gaps 2;
QY 281 TSAGDGVYVPEEFPPKVRITDAHMGILTDDETSIGRSHTSLVNDPGLNITVGVF 340
Db 24 SNAGDGVYVPEESLSVNHVLDHMGVLDGDDATGLRSHTSLVNDPGLNITVGVF 83
QY 341 ARRGYNISLAVGHAETKGISRTITVPATDESVKLYQOLKLYDVHEVHL--THLP 397
Db 84 ARRGYNISLAVGPAEKESISRTITVPETVESIEKXVASSTSLMCNESSMTLEPHL- 142
QY 398 FSEBELMKIIVNNAARDVDIASIFRAKAVDVSHTITLQLTGDDKMAQLRL 455
Db 143 ILKGNLSISKVSVNTAARKEILHIAQIFRAKPVDSHTVTLQLTGDDKMAVDTKVI 200
RESULT 10
ABB08418
ID ABB08418 standard; protein; 133 AA.
XX ABB08418;
AC
XX
XX 18-JUN-2002 (first entry)
DT
XX
XX Wheat ALS small subunit (clone wdk2c.pk015.a13).
DE
XX
XX ALS; Acetolactate synthase; acetoxyhydroxy synthase;
KM amino acid biosynthesis; plant; herbicide; ALS small subunit;
KW ALS holoenzyme; crop protection chemical; enzyme.
XX
XX Triticum aestivum.
OS
XX
XX US2001044939-A1.
PN
XX
XX 22-NOV-2001.
PD
XX
XX 08-DEC-2000; 2000US-0132618.
PF
XX
XX 04-JAN-2000; 2000US-174437P.
PR
XX
XX (ABELL/) ABELL L M.
PA (FALCO/) FALCO S C.
PA (FAMO/) FAMODU O O.

XX Abell LM, Falco SC, Famodu OO;
PI WPI: 2002-113395/15.
XX N-PSDB; ABA98822.
DR
XX New small subunits of plant acetolactate synthase and nucleic acids
PT encoding them, useful in screening for novel crop protection chemicals
XX or potential herbicidal compounds based upon holoenzyme inhibition
PS Claim 19; Page 40; 57pp; English.
XX The invention relates to new isolated polypeptides, which comprises an
CC acetolactate synthase (ALS) small subunit. Acetolactate synthase is the
CC first committed step in branched chain amino acid biosynthesis in plants
CC and bacteria. The polypeptides are useful in aiding the discovery of new
CC herbicides that inhibit plant acetolactate synthase activity. The
CC polypeptides are useful for preparing plant ALS holoenzyme, which are
CC useful in screening for potential herbicidal compounds based upon
CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
CC for novel crop protection chemicals. Particularly, the polypeptides can
CC be used as a target to facilitate design and/or identification of
CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
CC polypeptides could be appropriate for new herbicide discovery and
CC design. The polynucleotides are also useful for producing the
CC polypeptides and for screening for new herbicides. The current sequence
CC represents the wheat ALS small subunit (clone wdk2c.pk015.a13).
XX Sequence 133 AA;
SQ
Query Match 20.4%; Score 499; DB 23; Length 133;
Best Local Similarity 73.3%; Pred. No. 9.6e-39;
Matches 96; Conservative 22; Mismatches 13; Indels 0; Gaps 0;
QY 108 ARRGYNISLAVGNLNDKALFTIIVCGTERVQYVIEDLQKLVNWKVEDISSEPOVERE 167
Db 3 ARRGYNIESLAVGNLNDKALFTIIVSGTDRVAKQVIEQLNKLNVNLNVEDLSKEPOVERE 62
QY 168 LMLVNVNAHPEERAEFMVLVDFFRAVVDIAERHALTIEVTGPGKMIANERLKKFOIRE 227
Db 63 LMLIKLNPEEDRADVDFVAVNFRKAVVDISENSLLEVTGPGKIVAAQRNLRRFGIEE 122
QY 228 IVRTGRIALRR 238
Db 123 ICRGTGRIALRRQ 133
RESULT 11
ABB08409
ID ABB08409 standard; protein; 156 AA.
XX ABB08409;
AC
XX
XX 18-JUN-2002 (first entry)
DT
XX
XX Corn ALS small subunit (clone cen3n.pk0112.c11 - EST fragment).
DE
XX
XX ALS; Acetolactate synthase; EC 4.1.3.18; acetoxyhydroxy synthase;
KM amino acid biosynthesis; plant; herbicide; ALS small subunit;
KW ALS holoenzyme; crop protection chemical; enzyme.
XX
XX Zea mays.
OS
XX
XX
XX Key Location/Qualifiers
FH
FT Misc-difference 35
FT //label= unknown
FT //note= "Encoded by GNN"
FT
FT Misc-difference 43
FT //label= unknown
FT //note= "Encoded by GNC"
FT
FT Misc-difference 54
FT //note= "Encoded by GTN"
FT
FT Misc-difference 55

FT /label= unknown
 FT /note= "Encoded by GNT"
 FT Misc-difference 57
 FT /label= unknown
 FT /note= "Encoded by ANA"
 FT Misc-difference 58
 FT /label= unknown
 FT /note= "Encoded by NCT"
 FT Misc-difference 64
 FT /label= unknown
 FT /note= "Encoded by ANA"
 FT Misc-difference 66
 FT /label= unknown
 FT /note= "Encoded by GNA"
 FT Misc-difference 95
 FT /note= "Encoded by GTN"
 FT Misc-difference 98
 FT /note= "Encoded by GTN"
 FT Misc-difference 102
 FT /note= "Encoded by GTN"
 FT Misc-difference 119
 FT /label= unknown
 FT /note= "Encoded by ANA"
 FT Misc-difference 130
 FT /label= unknown
 FT /note= "Encoded by GNA"
 FT Misc-difference 149
 FT /label= unknown
 FT /note= "Encoded by NAT"
 FT Misc-difference 150
 FT /note= "Encoded by CCN"
 FT Misc-difference 151
 FT /label= unknown
 FT /note= "Encoded by GNG"
 PN US200104939-A1.
 XX
 XX 22-NOV-2001.
 PD
 XX 08-DEC-2000; 2000US-0732618.
 PF
 XX 04-JAN-2000; 2000US-174437P.
 PR
 XX (ABELL) ABELL L M.
 PA (FALC/) FALCO S C.
 PA (FAMO/) FAMODU O O.
 XX
 PI Abell LM, Falco SC, Famodu OO;
 XX
 DR WPI; 2002-113395/15.
 DR N-PSDB; ABA98813.
 XX
 PT New small subunits of plant acetolactate synthase and nucleic acids
 PT encoding them, useful in screening for novel crop protection chemicals
 PT or potential herbicidal compounds based upon holoenzyme inhibition
 XX
 PS Claim 19; Page 27-28; 57pp; English.
 XX
 CC The invention relates to new isolated polypeptides, which comprises an
 CC acetolactate synthase (ALS) small subunit. Acetolactate synthase is the
 CC first committed step in branched chain amino acid biosynthesis in plants
 CC and bacteria. The polypeptides are useful in aiding the discovery of new
 CC herbicides that inhibit plant acetolactate synthase activity. The
 CC polypeptides are useful for preparing plant ALS holoenzyme, which are
 CC useful in screening for potential herbicidal compounds based upon
 CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
 CC for novel crop protection chemicals. Particularly, the polypeptides can
 CC be used as a target to facilitate design and/or identification of
 CC inhibitors of those enzymes that may be useful as herbicides. Thus, the
 CC polypeptides could be appropriate for new herbicide discovery and
 CC design. The polynucleotides are also useful for producing the
 CC polypeptides and for screening for new herbicides. The current sequence
 CC represents the corn ALS small subunit (clone cen3n.pk0112.c11) encoded

CC by an EST fragment.
 XX
 SQ Sequence 156 AA;
 Query Match 19.3%; Score 470.5; DB 23; Length 156;
 Best Local Similarity 63.5%; Pred. No. 5.7e-36;
 Matches 99; Conservative 15; Mismatches 37; Indels 5; Gaps 2;
 QY 141 OYIEOLKLVNLYKVEDISEPOVERELMLVNVNAPESRAEIMLVDFRVRVYDIAEH 200
 DB 1 QYIEOLKLVNLYKVEDISEPOVERELMLVNVNAPESRAEIMLVDFRVRVYDIAEH 60
 QY 201 ALTEVTGDPGKMIANVERNLKRFQIREIVRTGKIALRREKMGATAPFMRSAASYPDLKE 260
 DB 61 TLTXEXAGDPGKIAAYQRLNRKRFQIREICRTGKIALRREKMGATAPFMRSAASYPDLKE 120
 QY 261 QAPVSVLRSSSKR---GAIVPQKETSAGGVYVEEPP 293
 DB 121 ALPKRPLTSMKTVNGSFV--RPSNAGDVPXPESY 154
 RESULT 12
 ABB08413
 ID ABB08413 standard; protein; 156 AA.
 XX
 AC ABB08413;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Rice ALS small subunit (clone r10.pk084.a24 - EST fragment).
 XX
 KW ALS; Acetolactate synthase; EC 4.1.3.18; acetylhydroxy synthase;
 KW amino acid biosynthesis; plant; herbicide; ALS small subunit;
 KW ALS holoenzyme; crop protection chemical; enzyme.
 XX
 OS Oryza sativa.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 154
 FT /note= "Encoded by CCN"
 FT
 PN US200104939-A1.
 XX
 XX 22-NOV-2001.
 PD
 XX 08-DEC-2000; 2000US-0732618.
 PF
 XX 04-JAN-2000; 2000US-174437P.
 PR
 XX (ABELL) ABELL L M.
 PA (FALC/) FALCO S C.
 PA (FAMO/) FAMODU O O.
 XX
 PI Abell LM, Falco SC, Famodu OO;
 XX
 DR WPI; 2002-113395/15.
 DR N-PSDB; ABA98817.
 XX
 PT New small subunits of plant acetolactate synthase and nucleic acids
 PT encoding them, useful in screening for novel crop protection chemicals
 PT or potential herbicidal compounds based upon holoenzyme inhibition
 XX
 PS Claim 19; Page 33-34; 57pp; English.
 XX
 CC The invention relates to new isolated polypeptides, which comprises an
 CC acetolactate synthase (ALS) small subunit. Acetolactate synthase is the
 CC first committed step in branched chain amino acid biosynthesis in plants
 CC and bacteria. The polypeptides are useful in aiding the discovery of new
 CC herbicides that inhibit plant acetolactate synthase activity. The
 CC polypeptides are useful for preparing plant ALS holoenzyme, which are
 CC useful in screening for potential herbicidal compounds based upon
 CC holoenzyme inhibition. The ALS holoenzyme is also useful in screening
 CC for novel crop protection chemicals. Particularly, the polypeptides can

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144881.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.

PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 13.6%; Score 332; DB 21; Length 86;

Best Local Similarity 77.9%; Pred.No. 2,3e-23; Indels 0; Gaps 0;

Matches 67; Conservative 7; Mismatches 12;

QY 404 MLKIVNMAARDVDIDIASIFRAKAVDSDHITLQLTGDDKMWALQRLPEYICIEV 463
DB 1 MLKIVADTSARDVDIDIQVFRAKAIDVSDHITILEVGDLRKKSALOTQLEAVGICIEV 60
QY 464 ARTGRVALARESGVDSKYRGYSFLI 489
DB 61 ARTGRVALVRESGVDSYLRGYSPL 86

Search completed: March 10, 2003, 09:07:58
Job time : 43 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 09:06:30 ; Search time 21 Seconds
(without alignments)
2247.714 Million cell updates/sec

Title: US-09-997-900-2
2444

Perfect score: 1 MAAISVSSPSIRCLRACS.....ARESGVDSKYLGRYFLLTG 491

Sequence: 1 MAAISVSSPSIRCLRACS.....ARESGVDSKYLGRYFLLTG 491

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2358.5	96.5	484	2 D84725	probable acetolact
2	463	18.9	182	2 AC3884	acetolactate synth
3	461	18.9	188	2 S74358	acetolactate synth
4	427	17.5	174	2 S73151	acetohydroxyacid s
5	359.5	14.7	159	2 F69464	acetolactate synth
6	359.5	14.7	192	2 E70459	acetolactate synth
7	351	14.4	168	2 E70855	probable ilvN prot
8	345	14.1	174	2 T35829	acetolactate synth
9	341	14.0	169	2 T45414	hypothetical prote
10	329.5	13.5	174	2 E69644	acetolactate synth
11	327.5	13.4	165	2 F97290	acetolactate synth
12	327.5	13.4	168	2 B69059	acetolactate synth
13	327	13.4	175	2 JC7166	acetolactate synth
14	318.5	13.0	173	2 D84032	acetolactate synth
15	315	12.9	163	2 H81066	acetolactate synth
16	307.5	12.6	172	2 B64320	acetolactate synth
17	307	12.6	172	2 B56684	acetohydroxy acid
18	305	12.5	166	2 B97922	acetolactate synth
19	302.5	12.4	158	2 F95051	acetolactate synth
20	302	12.4	167	2 JC5165	acetolactate synth
21	290.5	11.9	171	2 C72362	acetolactate synth
22	280	11.5	190	2 AD3329	acetolactate synth
23	277	11.3	163	2 F83059	acetolactate synth
24	277	11.3	185	2 I40667	acetolactate synth
25	276.5	11.3	163	1 YCEC3H	acetohydroxy acid
26	275	11.3	190	2 D97604	acetolactate synth
27	275	11.3	190	2 AE2826	acetohydroxy acid
28	273.5	11.2	164	2 D82072	acetolactate synth
29	271.5	11.1	163	2 S15940	acetolactate synth

30	271.5	11.1	164	2 AB0517	acetolactate synth
31	268.5	11.0	163	2 B64131	acetolactate synth
32	268.5	11.0	163	2 B90639	acetolactate synth
33	268.5	11.0	163	2 B85490	acetolactate synth
34	268.5	11.0	164	2 AB0067	acetolactate synth
35	263.5	10.8	163	2 AB1694	acetolactate synth
36	263.5	10.8	163	2 AI1322	acetolactate synth
37	262.5	10.7	158	2 AB6778	acetolactate synth
38	260.5	10.7	158	2 S31139	hypothetical prote
39	255	10.4	142	2 B48648	probable acetolact
40	252.5	10.3	158	2 E84956	acetohydroxy acid
41	251	10.3	232	2 T39432	acetolactate synth
42	235.5	9.6	309	2 S19411	acetolactate synth
43	223.5	9.1	330	2 T49352	probable acetolact
44	208.5	8.5	154	2 F81404	acetolactate synth
45	118	4.8	2245	2 T27511	hypothetical prote

ALIGNMENTS

RESULT 1	
D84725	probable acetolactate synthase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)	
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001	
C:Accession: D84725	
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y. M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter Nature 402, 761-768, 1999	
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.	
A:Reference number: AB4420; MUID:20083487; PMID:10617197	
A:Accession: D84725	
A:Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-484 <STO>	
A:Cross-references: GB:AE002093; NID:g4887755; PIDN:AA032291.1; GSPDB:GN00139	
C:Genetics:	
A:Gene: At2g31810	
A:Map position: 2	
Query Match	96.5%; Score 2358.5; DB 2; Length 484;
Best Local Similarity	97.2%; Pred. No. 5-5e-150;
Matches 480; Conservative 0; Mismatches 1; Indels 13; Gaps 2;	
QY 1 MAAISVSSPSIRCLRACSDDSPALVSTRVSPAKISYLSGISHRDGMKRMGEFV 60	
DB 1 MAAISVSSPSIRCLRACSDSPALVSTRVSPAKISYLSGISHRDGMKRMGEFV 60	
QY 61 RSYDGRISDASFSSEASAPPKSVKRNHTISVPYGDSSGMINRIAGYFARGNIESLANG 120	
DB 61 RSYDGRISDASFSSEASAPPKSVKRNHTISVPYGDSSGMINRIAGYFARGNIESLANG 120	
QY 121 LNRDKALFTIVVCGTFRVLOQVTEQLKLVNVLKVEDISEPQVEKMLVKNAPESR 180	
DB 121 LNRDKALFTIVVCGTFRVLOQVTEQLKLVNVLKVEDISEPQVEKMLVKNAPESR 180	
QY 181 AETIMLVDTFRARVVDIAEHALTIEVTGPGKMI AVERNMLKFOIREIVRTGK---IALR 237	
DB 181 AETIMLVDTFRARVVDIAEHALTIEVTGPGKMI AVERNMLKFOIREIVRTGK---IALR 237	
QY 238 REEMGATAPFWRPSAASYPDLKEQAPVSVLRSSKKAIVPQKETSAGSGGVYVPEPFDFK 297	
DB 241 REEMGATAPFWRPSAASYPDLKEQAPVSVLRSSKKAIVPQKETSAGSGGVYVPEPFDFK 300	
QY 298 VHRILDAHNGLLTDEDTSGLRSHHTLSILVNDIPGVNIYTGAFARGVYINQSLAAGHAE 357	
DB 301 VHRILDAHNGLLTDEDTSGLRSHHTLSILVNDIPGVNIYTG-----SLAVGHAE 350	
QY 358 KGISRTITVYIPATDESVSRLVQOLYKLVVDVHEVHDLTHLPFSEREIMLIKIAVNAARD 417	
DB 351 KGISRTITVYIPATDESVSRLVQOLYKLVVDVHEVHDLTHLPFSEREIMLIKIAVNAARD 410	

QY 418 VLDIASIFRAKADVSDHTITLQDGLDKMVALQRLPEYGCICEVARTGRVALARESGV 477
|||||
Db 411 VLDIASIFRAKADVSDHTITLQDGLDKMVALQRLPEYGCICEVARTGRVALARESGV 470

QY 478 DSKYLRGYSFLLTG 491
|||||
Db 471 DSKYLRGYSFLLTG 484

RESULT 2
AC2384
acetylactate synthase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AC2384
R:Kaneoka, T.; Nakamura, Y.; Wolke, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AC2384
A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-182 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA076326.1; PID:g17133764; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: llvN

C:Superfamily: acetylactate synthase small chain

Query Match
Best Local Similarity 18.9%; Score 463; DB 2; Length 182;
Matches 95; Conservative 29; Mismatches 30; Indels 2; Gaps 1;

QY 86 KHTISVVGDSGMINRIAGVFARAGYNIESLAVGLNRDQAL--FTIYVCGEYVLAQYI 143
|||||
Db 11 KHTLSLVDEDEGVALSRISLFAARGFNIESLAVSGESGVSRTIMVPGDRIEQLT 70

QY 144 EQLQKLVNVLKVEDISSEPOVERELMLVNVNAHPESRAELMLVDTFRARVVDIAEHALT 203
|||||
Db 71 KOLYKLVNVLKVDITETPCVERELMLKVNATSSNRSEVIELAQIFRRARVVDIAEDSLT 130

QY 204 IEVTGDPGKMAIERNLKKFOIREIVRTGKIALRRE 239
|||
Db 131 LEVYDGPCKMAIYOVYQKFGIREIAKTALTR 166

RESULT 3

574358
acetylactate synthase llvN - Synecchocystis sp. (strain PCC 6803)

N:Alternate names: hypothetical protein sl10065

C:Species: Synecchocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

C:Accession: 574358

R:Kaneoka, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.
DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis sp.

A:Reference number: 574322; MUID:97061201; PMID:8905231

A:Accession: 574358

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-188 <KAN>

A:Cross-references: EMBL:D64001; GB:AB001339; NID:g1001102; PIDN:BA10276.1; PID:g100113

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Gene: llvN

A:Start codon: GTG

C:Superfamily: acetylactate synthase small chain

Query Match
Best Local Similarity 18.9%; Score 461; DB 2; Length 188;
Matches 87; Conservative 40; Mismatches 37; Indels 0; Gaps 0;

QY 320 HTLSLVNDIPGVNLIVTGVFARAGYNISLAVGAEFGISRTITVIPATDESYSKLQV 379
|||||
Db 19 HTLSLVDEDEGVALSRISLFAARGFNIESLAVSGESGVSRTIMVPGDRIEQLT 78

QY 380 QLYKLVNVEHVDLTLFPERELMLKIVANNAARDVLDIASIFRAKADVSDHTITL 439
|||||
Db 79 QLYKLVNVLKVDITETPCVERELMLKVNATSSNRSEVIELAQIFRRARVVDIAEDSLT 138

QY 440 QLTGDDKMAVALQRLPEYGCICEVARTGRVALARESGVDSKYL 483
|||
Db 139 EYVGDGPKMAIYOVYQKFGIREIAKTALTR 182

RESULT 4

573151
acetylhydroxyacid synthase small chain - red alga (Porphyra purpurea) chloroplast

C:Species: chloroplast Porphyra purpurea

C:Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 22-Jun-1999

C:Accession: 573151

R:Reith, M.; Munholland, J.

Plant Mol. Biol. Rep. 13, 333-335, 1995

A:Title: Complete nucleotide sequence of the Porphyra purpurea chloroplast genome.

A:Reference number: 573108

A:Accession: 573151

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-174 <REI>

A:Cross-references: EMBL:U38804; NID:g1276652; PIDN:AA08116.1; PID:g1276696

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995

C:Genetics:

A:Gene: chlH

C:Superfamily: acetylactate synthase small chain

C:Keywords: chloroplast

Query Match
Best Local Similarity 17.5%; Score 427; DB 2; Length 174;
Matches 83; Conservative 41; Mismatches 40; Indels 0; Gaps 0;

QY 320 HTLSLVNDIPGVNLIVTGVFARAGYNISLAVGAEFGISRTITVIPATDESYSKLQV 379
|||||
Db 3 HTLSLVDEDEGVALSRISLFAARGFNIESLAVSGESGVSRTIMVPGDRIEQLT 62

QY 380 QLYKLVNVEHVDLTLFPERELMLKIVANNAARDVLDIASIFRAKADVSDHTITL 439
|||||
Db 63 QLYKLVNVLKVDITETPCVERELMLKVNATSSNRSEVIELAQIFRRARVVDIAEDSLT 122

QY 440 QLTGDDKMAVALQRLPEYGCICEVARTGRVALARESGVDSKYL 483
|||
Db 123 EYVGDGPKMAIYOVYQKFGIREIAKTALTR 166

RESULT 5

569464
acetylactate synthase, small subunit (llvN) homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Jun-1999

C:Accession: 569464

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E.

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Arlacher, P.; Kaine, B.P.; Sykes, Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: 569464

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA
A.Residues: 1-159 <KLE>
A.Cross-references: GB:AE000985; GB:AE000782; NID:g2689308; PIDN:AAB9532.1; PID:g264883
C:Superfamily: acetolactate synthase small chain

Query Match 14.7%; Score 359.5; DB 2; Length 159;
Best Local Similarity 48.4%; Pred. No. 4.1e-17;
Matches 75; Conservative 36; Mismatches 41; Indels 3; Gaps 2;

QY 86 KHTISVFGDESGMIRINGVFARRGYNIESLAVGLNR--DKALFTIYVCGTEFVQOYI 143
DB 2 KHTIALVFNKRGVLAARVAGLFRRRGFINIESLVTGTEDDDSRMTIYVGGDKVVEQYI 61
QY 144 EOLQKLVNLYKVEDISSPOVERELMLVKNNAHPESRAEIMLVDTFRARVYDIAEHALT 203
DB 62 KQNKLIKETIKVSEI-TESSVERELCLIVHAPEKRGELVELTNIFRARIYVSDSFI 120
QY 204 IEVTGDPGKMAVERNLKFKQIREIVRTGKIALRR 238
DB 121 IEVTGDEKVSFAIDLMROYGKELARTGKVAMVR 155

RESULT 6

E70459
acetolactate synthase - Aquifex aeolicus

C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 22-Jun-1999
C:Accession: E70459

R.Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O.V.

Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: E70459
A:Residues: 1-192 <AQC>
A:Cross-references: GB:AE000759; NID:g2984125; PIDN:AAC07662.1; PID:g2984126; GB:AE00065
A:Experimental source: strain VFS
C:Genetics:
A:Gene: llvH
C:Superfamily: acetolactate synthase small chain

Query Match 14.7%; Score 359.5; DB 2; Length 192;
Best Local Similarity 47.9%; Pred. No. 5.3e-17;
Matches 78; Conservative 35; Mismatches 45; Indels 5; Gaps 3;

QY 81 KSKVRRHTISVFGDESGMINRIAGVPARRGVIESLANGLNDRKAL--FTIYVCGTERV 138
DB 22 KGVRRKHITIVKVRNEMGVLAIRATLACRGVIEGTSVGETHEKGISRMTIEVIGDIY 81
QY 139 LOQVIOLOKLVNLYKVEDISSPOVERELMLVKNNAHPESRA--ETIMLVDTFRARVYD 196
DB 82 IEVYQQLRLRLDITLVNLDITLVDPHVERELALTKVYT--PSSRADEVLRTITELRGKYVD 140
QY 197 IAEHALTIEVTGDPGKMAVERNLKFKQIREIVRTGKIALRR 239
DB 141 VSPDFTYIEVTGDEKINMIELKPKGIREMARITGKVAMRRE 183

RESULT 7

E70855
probable llvN protein - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: E70855

R.Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: E70855
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-168 <COI>

A:Cross-references: GB:AL021287; GB:AL123456; NID:g3261508; PIDN:CA16087.1; PID:g2729
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: llvN
C:Superfamily: acetolactate synthase small chain

Query Match 14.4%; Score 351; DB 2; Length 168;
Best Local Similarity 44.0%; Pred. No. 1.6e-16;
Matches 73; Conservative 40; Mismatches 47; Indels 6; Gaps 2;

QY 79 TPFSKVRKHTISVFGDESGMIRINGVFARRGYNIESLAVGLNR--RDALFTIYVCGTE 136
DB 2 SPKT-----HTLSVLDKRGVLAARVALLSRGFINIESLAVGATECKDSRMTIYVSAED 57
QY 137 RVLOQVTEOLOKLVNLYKVEDISSPOVERELMLVKNNAHPESRAEIMLVDTFRARVYD 196
DB 58 TPLEQITQKQNLKLVNLYKIVEDODEHSVERELALIKVQADAGSRQVTEFVNLFRANVID 117
QY 197 IAEHALTIEVTGDPGKMAVERNLKFKQIREIVRTGKIALRRKMG 242
DB 118 VSPESLTVEATGNRKLALRLVLEPFGIREIAOSGWSLSRQPG 163

RESULT 8

T35829
acetolactate synthase small subunit - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
C:Accession: T35829

R.Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1999

A:Reference number: Z21589
A:Accession: T35829
A>Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A:Residues: 1-174 <MUR>
A:Cross-references: EMBL:AL035566; PIDN:CAB37589.1; GSPDB:GN00070; SCOEDB:SC8D9.25
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC8D9.25
C:Superfamily: acetolactate synthase small chain

Query Match 14.1%; Score 345; DB 2; Length 174;
Best Local Similarity 43.9%; Pred. No. 4.3e-16;
Matches 69; Conservative 36; Mismatches 52; Indels 0; Gaps 0;

QY 317 LRSHTLSLVNDIPGVNITVGVFARRGVNIOSLAVGAEHTKGISRTIYIPATDESYSK 376
DB 1 MSKHTLSVLENTPGVLAIRATLALSRGFNIDSLAVGVTMHPDISRTIYVSEDPLEQ 60
QY 377 LVQOLYKLVNLYKVEDISSPOVERELMLIKIVNNAARDVDIASIPRAKAVDSDHT 436
DB 61 VTRQKLVNLYKVELEPQAVQRELVLKVRSDMETSQIYEIQVLPRAKIVDVSPA 120
QY 437 ITIQLTGDLDMVALORLLEPGICVARTGRVALAR 473
DB 121 VITEATGSSDKLTAMLMLEPGIKELVOSGTIAIGR 157

RESULT 9

T45414
hypothetical protein [imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 18-Feb-2000

C:Accession: T45414

R.Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1997

A:Reference number: Z16918
A:Accession: T45414

THIS PAGE BLANK (USPTO)